Spring 2020

Elementary Physical Education Teacher Perceptions of Motor Skill Assessment

Jenna Fisher

ELEMENTARY PHYSICAL EDUCATION TEACHER PERCEPTIONS OF MOTOR SKILL ASSESSMENT

by

Jenna Fisher

Bachelor of Science
East Stroudsburg University of Pennsylvania, 2010

Master of Science
East Stroudsburg University of Pennsylvania, 2012

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Physical Education

College of Education

University of South Carolina

2020

Accepted by:

Ali Brian, Major Professor

Jacalyn Lund, Committee Member

Collin A. Webster, Committee Member

David F. Stodden, Committee Member

Danielle Nesbitt, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

DEDICATION

Dedicated to the future of assessment in physical education.

ACKNOWLEDGEMENT

Dr. Ali Brian- Thank you for recognizing my passion for learning and offering to advise me. I appreciate your efforts to gently and gradually show me how to be a more self-sufficient scholar and researcher. Thank you for being a sensitive and patient advisor who views doctoral students as growing young scholars in need of developmentally appropriate learning experiences.

Dr. Jacalyn Lund- I really appreciate you for agreeing to become part of my doctoral journey, twice. I cannot thank you enough for your hard work regarding assessment in physical education; it sparked my interest long ago and has fueled a fire in me to this day. I have so much more to learn from you. Thank you!

Drs. Webster, Stodden, & Nesbitt- Thank you so much for your continued support through my many iterations of a dissertation. I value your ideas, feedback, and support.

UofSC Graduate Assistants- Thank each and every one of you for acting as the best support system throughout this journey, for having my back along the way, and for teaching me things that I never thought I could learn. Cheers to being lifelong friends!

Friends & Family- Thank you so much for your continued support and understanding of my need to temporarily hide under a rock for three years. I love you all!

My husband, Ryan Montero- Without your support, I would not be here. I am so thankful for everything you have sacrificed for me to be able to accomplish this feat. This is not only my accomplishment, but yours as well. Cheers to a long healthy life together. I love you. Thank you for everything.

ABSTRACT

Though a critical aspect of the teaching-learning process in physical education, the assessment of student learning in physical education has received little investigative attention. In line with the primary focus on motor competence in the United States National K-12 Physical Education Standards, the purpose of this dissertation is to explore elementary physical education teachers' assessment behaviors with respect to assessing students' motor skills. This dissertation consists of three studies. In the first study, a survey was developed using the major themes from existing physical education assessment literature and then tested for content validity via the modified Delphi method. The second study involved examining the psychometric properties of one of the motor skill assessment behavior survey subscales, focusing on the perception of motor skill assessment using a sample of current inservice elementary physical education teachers across the United States of America. The third study reports the descriptive and predictive statistics from Study 2. The results of this dissertation provide key information related to elementary physical education teachers' perceptions and perceived needs regarding current nationwide motor skill assessment behaviors.

Keywords: Formal accountability, standards-based, student learning, survey, pedagogy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EFA ..... Exploratory Factor Analysis

E.S.S.A. ..... Every Student Succeeds Act

SHAPE America ..... Society of Health and Physical Educators

CHAPTER 1

INTRODUCTION

This dissertation consists of three studies that examine elementary physical education teachers' motor skill assessment behavior. The first study consists of survey item development and established content validity of the 'Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey' using a sample of physical education assessment content experts and inservice elementary physical education teachers. The second study examines the psychometric properties (internal consistency reliability and factor structure) of the 'Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey Subscale' (a survey subscale from the larger survey in Study 1) with a national sample of elementary physical education teachers across the United States of America. The third study is a secondary data analysis examining the descriptive statistics related to participant responses to the 'Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey Subscale', group differences based on motor skill assessment behavior (assess motor skills versus do not assess motor skills), and predictive statistics to predict future motor skill assessment behavior using responses to the perception subscale. This introductory chapter will provide an overarching description of this dissertation.

Background

Assessment in Physical Education

Formal assessment is described as the collection of evidence regarding student learning (Black & Wiliam, 2010; Emmanouilidou, Derri, Aggelousis, & Vassiliadou, 2012; Frapwell, 2010; Hay, 2006; Lund, 1992; Matanin & Tannehill, 1994; Melograno, 1997; Ni Chronin & Cosgrave, 2013; Pickup & Price, 2007; Piotrowski, 2000; Rink, 1993; Siedentop & Tannehill, 2000; Wright & van der Mars, 2004). Formal assessment is often used for the tracking, documenting, and reporting of student progress (Doolittle, 1996; Hay & Penney, 2009; Lund & Kirk, 2010; Mintah, 2003; Safrit, 1986; van der Mars, Timken, & McNamee, 2018; Veal, 1988). The evidence gleaned from student assessment can be used to demonstrate quality and effectiveness of an instructional unit or an entire physical education program (Bailey, 2001; Veal, 1988; Wiliam, 2011). When used systematically, objectively (Veal, 1988), and continuously throughout the teaching-learning process (Tousignant & Siedentop, 1983), assessment evidence informs educators of student learning needs (Black & Wiliam, 1988; Desrosiers, Genet-Volet, & Godbout, 1997; Safrit, 1986; Veal, 1988). An educator should therefore improve his or her instructional and pedagogical practices based on student assessment evidence to better facilitate student learning (Veal, 1988; Wiliam, Lee, Harrison, & Black, 2004).

Student assessment is one of the four main pillars making up the Essential Components of Physical Education (Society of Health and Physical Educators [SHAPE America], 2015). The Essential Components of Physical Education Guidance Document is comprised of four essential components in physical education dedicated to providing structure for quality physical education programming. Along with student assessment,

SHAPE America (2015) makes recommendations for policy and environment, curriculum, and appropriate instruction in physical education. Within the student assessment component, quality physical educators are encouraged to formally collect evidence of student learning continually throughout the learning cycle by using a variety of different assessment methods that are directly aligned with national and/or state physical education standards and SHAPE America grade-level outcomes. Consequently, as recommended by SHAPE America, the national organization for physical education teachers in the United States of America, physical educators should utilize assessment within their programs.

Additionally, certified physical education teachers who graduate from accredited universities in the United States of America should possess the skills necessary to formally collect evidence of student learning via assessment (SHAPE America, 2017). The SHAPE America 2017 Initial Physical Education Teacher Education Standards state that "physical education candidates select and implement appropriate assessments to monitor students' progress and guide decision making related to instruction and learning" (SHAPE America, 2017, p. 4). Therefore, one can assume that physical education teachers currently working in the field who have graduated from an accredited program, are prepared to assess student learning and do so in accordance with the national and/or state standards and grade level outcomes.

Problems with Assessment in Physical Education

Lack of Accountability. Issues regarding assessment in physical education have been prevalent since the 1970's (Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013). Historically, physical education teachers have not been held

accountable at the state or national level to assess student learning (Rink, Jones, Kirby, Mitchell, & Doutis, 2007; van der Mars et al., 2018) with a notable temporary exception taking place in the state of South Carolina (See Rink & Mitchell, 2003). Without accountability measures in place, non-core subject areas traditionally have failed to formally assess program goals (Rink et al., 2007). To our knowledge, since the accountability era in the United States of America in the 1980's, approximately five research studies regarding physical education teacher assessment practices (frequency and context of assessment) were conducted in the United States of America (See Hensley, Lambert, Baumgartner, & Stillwell, 1987; Imwold, Rider, & Johnson, 1982; Kneer, 1986; van der Mars., et al 2018a, van der Mars et al., 2018b). An additional study measured preservice physical education teacher assessment practices (see Lund & Veal, 2008). The lack of assessment research in physical education provides evidence that assessment is not only omitted from the culture of school physical education (Lund & Veal, 2008) but is also lacking from the physical education research agenda. Moreover, the majority of research on assessment in physical education surrounds teachers' perceptions and is outdated (van der Mars et al., 2018) with near complete omission at the elementary school level (Imwold, Rider, & Johnson, 1982; Ni Chronin & Cosgrave, 2013).

Lack of Quality. Assessment practices are often neglected and avoided in physical education (Lander et al., 2016, 2017; Leirhaug & MacPhail, 2015). When assessment is used in physical education, it is often used inappropriately (Borghouts et al., 2017; Lund & Veal, 2008). Examples of inappropriate use of include teachers assessing based on subjective criteria to measure student performances, only using summative assessments (Borghouts et al., 2017; Hay & Macdonald, 2008; Lund & Veal,

2008), and only formally assessing managerial tasks (i.e., changing clothes for physical education and attendance) as opposed to learning content (van der Mars et al., 2018). A likely cause for assessment misuse stems from the major personal barriers toward the use of assessment, including the teachers' perceived lack of knowledge about assessment and how to use it (Kneer, 1986; Lander et al., 2016, 2017; Matanin & Tannehill, 1994; Veal, 1988). Lander et al. (2015) reports that a physical education teachers' perception about assessment likely influences his or her actual assessment behavior.

Perceived Barriers. Barriers in the assessment process are often the cause for teachers' avoidance of assessment in physical education (Morgan & Hansen, 2008; Penney, 2012; Stiggins, 1997) and a likely cause in teachers' misuse of assessment stems from personal barriers toward using it, including teachers' perceived lack of knowledge about assessment and how to use assessment (Kneer, 1986; Lander et al., 2016, 2017; Matanin & Tannehill, 1994; Veal, 1988). Assessment is considered one of the most difficult tasks in physical education (Morgan & Hansen, 2007). In combination with low accountability expectations for physical education assessment (van der Mars et al., 2018), Lander et al. (2015) reports that a physical education teachers' perceptions about assessment likely influences their actual assessment behavior. Physical education teachers often perceive that assessments that are too time consuming, require large amounts of equipment and set-up, have too many students with too few assessment administrators, take place in unauthentic settings, and require unattainable equipment and facilities (DinanThompson & Penney, 2015; Georgakis, Wilson, & Evans, 2015; Kneer, 1986; Lander et al., 2015, 2016, 2017; Leirhaug & MacPhail, 2015; MacPhail & Halbert, 2010; Michael et al., 2016; Mintah, 2003; Morgan & Hansen, 2007; Veal, 1988; Wiart &

Darrah, 2001). The most detrimental perceived barriers of assessment practices include a lack of understanding for the purpose of assessment, low perceived value and benefit of assessment, and a lack of training for how it should be used (Annerstedt & Larsson, 2010; Lander et al., 2015, 2016, 2017; Lund & Veal, 2008; Michael et al., 2016; Slingerland et al., 2017; van der Mars et al., 2018). Teacher buy-in and assessment philosophy presented major barriers regarding the use of standards-based assessment in California teachers (Michael et al., 2016). With a lack of accountability, district administrative policies, curriculum, and support, physical education teachers often forgo the assessment process altogether (Lander et al., 2015; Michael et al., 2016; van der Mars et al., 2018). Physical education teachers perceive a need for authentic assessments that can be completed in a natural physical education setting (Haynes & Miller, 2015; James, Griffin, & Dodds, 2009; Lander et al., 2016; MacPhail & Halbert, 2010; Mintah, 2003; Ni Chroinin & Cosgrave, 2013; Patton & Griffin, 2008; Richard, Godbout, Tousignant, & Grehaigne, 1999). Physical education teachers also have a strong desire for assessment training to improve their perceived inadequacies for assessment (Annerstedt & Larsson, 2010; Lander et al., 2015, 2017; Lund & Veal, 2008; Michael et al., 2016; Slingerland et al., 2017).

Motor Skills

SHAPE America's National Physical Education Standard 1 states "The physically literate individual demonstrates competency in a variety of motor skills and movement patterns" (SHAPE America, 2013). Motor skills (i.e., locomotor, manipulative, and non-locomotor skills) make up the majority of the content taught in physical education at the elementary level (see the Grade Level Outcomes for K-12 Physical Education; SHAPE

America, 2013), with goals of students mastering basic motor skills before the end of grade five (SHAPE America, 2013). An individual who is competent in motor skills at a young age is more likely to be able to perform more complex movements and be physically active as they grow older (Lima et al., 2017a, 2017b; Lubans, Morgan, Cliff, Barnett, & Okley, 2010). Therefore, in accordance with SHAPE America and recent motor behavior research, the assessment of student mastery of motor skills is most logical at the elementary level.

At present, physical education teacher practices and perceptions of assessment of students' motor skills remains unexplored. Currently, fitness testing is the most popular physical education content assessed in physical education (Lopez-Pastor et al., 2013). In 2018, the SHAPE America Research Council declared a national call for physical educators to be assessing motor skill competency in schools (Castelli & van der Mars, 2018) providing impetus to investigate physical education teacher motor skill assessment practices at the elementary school level.

Measuring Teacher Assessment Behavior

To effectively measure assessment practices and perceptions in physical education would require a measurement tool that produces valid and reliable results based on its sample (Bandalos, 2018). Currently, there are no measurement tools (i.e., surveys, etc.) that produce psychometrically sound results that specifically measure elementary physical education teachers' practices and perceptions regarding the assessment of students' motor skills. Therefore, this dissertation will serve to create and disseminate a survey that demonstrates acceptable psychometric properties that can be used to measure

motor skill assessment practices and perceptions of elementary physical education teachers across the United States of America.

Purpose and Research Questions

Study 1

The purpose of Study 1 was to develop a survey tool ("Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey"), that demonstrates content validity and is framed within major themes across physical education assessment literature. The survey development was guided by a panel of physical education assessment content experts and current inservice elementary physical education teachers. The Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey is comprised of several subscales, each measuring a different aspect of assessment behavior, including motor skill assessment perceptions, frequency and types used, resources used, preferred learning opportunities, and improvements to increase the likelihood of using assessment. This study used a descriptive-analytic modified Delphi Method research design which allows for the initial creation of survey items to be derived from a literature review (Avella, 2016). Once items were created, the experts and inservice teachers participated in several rounds (Hsu & Sandford, 2007) to anonymously rate the survey items regarding relevance and clarity while providing constructive feedback for the improvement of items (Rowe & Wright, 1999). Expert consensus on item relevance, clarity, and inclusion within the survey scale demonstrated content validity of survey items. The findings from this study helped to determine the survey items included within the survey subscales and subsequently used in Study 2.

Study 2

The purpose of this study was to determine the psychometric properties of one of the survey subscales, the 'Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey' subscale, which measures elementary physical education teacher perceptions of motor skill assessment. This tool was disseminated to a sample of current elementary physical education teachers across the United States of America. We explored the initial factor loadings and obtained psychometric properties using the perception subscale with our sample. The findings from this study informed our ability to accurately describe elementary physical education teacher perceptions of motor skill assessment in Study 3.

Study 3

The purpose of Study 3 was to describe elementary physical education teacher perceptions of motor skill assessment using the "Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey" subscale and to make predictions about how to influence motor skill assessment behavior in elementary physical education. This study used a descriptive-analytic research design as a secondary data analysis from Study 2 to obtain descriptive and predictive statistics. The findings from this research were used to report the perceptions of motor skill assessment by elementary physical education teachers across the United States of America.

Delimitations and Limitations

Delimitations

The participants in Study 1 included a purposive sample of content experts and a selection of current inservice elementary physical education teachers who have proven to

be successful and knowledgeable regarding assessment in physical education. We measured participant assessment expertise to ensure quality ratings of survey items during content validation. We determined the participants' (i.e., content experts and current inservice teachers) level of expertise based on their level of published research, assessment content knowledge, use of assessment in the field of physical education, highest level of completed education, teaching experience in K-12 schools, and their current place of employment. For Study 2 and 3, we recruited participants using a stratified random sampling of public schools across the United States of America and through popular social media platforms, including Facebook groups and Twitter.

The "Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey" and perception subscale are delimited to measuring assessment behaviors regarding SHAPE America National Physical Education Standard 1, motor skills, and/or each state's version of the motor skill standard.

Survey items are derived from major themes emerging from two comprehensive literature reviews on practices and perceptions of assessment in physical education (Fisher et al., in preparation) and through the modified Delphi process with physical education assessment content experts. Therefore, the perspectives and philosophies for assessment in physical education within the surveys are reflective of the modified Delphi participants and their advocacy for assessment in physical education.

In order to measure a nationwide sample of elementary physical education teacher perceptions of assessment, we decided to use a quantitative survey comprised of survey items influenced by variables measured in former physical education assessment literature (both qualitative and quantitative). However, within this survey, there are also

opportunities for the survey participants to leave open-ended responses and an option to participate in a qualitative interview later. The quantitative survey items use a Likert-type item response scale, as this is a very popular method in social science research used to measure non-cognitive item responses and used to differentiate between different respondent groups (Bandalos, 2018).

The measurement of the psychometric properties of the survey scores (Study 2) are employed to ensure quality and interpretability of the research findings within Study 3. Without sound psychometric properties, it would be inappropriate to interpret the findings regarding the perceptions of the elementary physical education teachers' assessment of motor skills (Bandalos, 2018).

Limitations

This dissertation may have limitations. Participants included in Study 1 (modified Delphi participants) are a purposive sample. We did our best to obtain a well-rounded sample of participants. Due to the nature of survey research, participants for Study 2 and 3 are a convenience sample and were incentivized to participate for entry into a raffle for a $50 Amazon gift card. Participants could only be included in the recruitment process if his/her district email address was available on the school district website for which he/she worked, if his/her school district was one selected from the random stratified sampling method, if he/she decided to participate in the survey as seen on the social media platforms (Facebook or Twitter), or if he/she was contacted by the state physical education representative or state SHAPE America executive. Finally, Hawaii hides teacher district email addresses from public access, thereby making it impossible to add teachers from this state to our stratified random sampling of teachers. However, Hawaii

teachers did have the opportunity to gain access through social media, state physical education representatives, and/or state SHAPE America executives. Study 2 sample demographics were not completed by all participants, and therefore, we do not have specific demographic data (sex, race, ethnicity, and years of teaching experience) for some participants.

We were unable to measure survey response rate. Due to the nature of trying to reach a large sample across the United States of America, we used several recruitment methodologies. Since our mixed recruitment methodologies (i.e., manual collection of email addresses, advocacy through state level physical education representatives, state level SHAPE America executives, and dissemination via social media platforms) were considered *posteriori,* we were unable to capture a survey response rate. Due to our recruitment method, we are unable to determine the exact number of people who received access to this survey and determine specifically where they gained access.

Our Study 2 sample was only large enough to run an exploratory factor analysis and obtain internal consistency reliability but did not have enough responses to run a confirmatory factor analysis. Additionally, within our exploratory factor analyses, several items demonstrated crossloading factor values thus demonstrating possibility for another factor structure. Future research should investigate other exploratory factor structures and confirmatory factor analysis with larger sample sizes.

Additionally, several items within our survey are self-report behaviors. Self-reported behavior is not an objective measure of behavior and therefore, participants have the opportunity to respond in socially desirable ways (Bandalos, 2018). However, this survey method was selected to ensure breadth rather than depth of responses to explore

generally what is happening in regard to motor skill assessment practices in the United States of America. The ability to measure assessment practices across the nation will provide evidence that can be utilized to plan for ways to advance the field forward.

Significance and Innovation

Significance

This dissertation is significant as it has the potential to obtain sound psychometric properties of results from a quickly administered survey tool using a sample of elementary physical education teachers across the United States of America. This is one of the most extensive studies to examine assessment practices and perceptions of physical education teachers by using a national sample. To our knowledge, no other study in physical education assessment research in the United States of America has gained a nationwide sample. The results from this dissertation can inform the SHAPE America Research Council regarding the extent to which elementary physical education teachers are assessing motor skills and their perceptions regarding the process of motor skill assessment. Future research that intends to measure physical education teachers' assessment practices and perceptions can utilize the surveys developed in these studies to measure across other national standards and content areas in physical education.

Innovation

This dissertation is the first to survey elementary physical education teachers across the United States of America regarding assessment practices and perceptions of the assessment of motor skills (or any national standard, for that matter) using a content valid and psychometrically sound survey tool. The participation of elementary physical education teachers across the United States of America increases the ability to generalize

the results in regard to motor skill assessment practices and perceptions of elementary physical education teachers across this country. Historically, research studies that have examined similar variables have used geographically limited samples (i.e., one state, one district). This dissertation is novel because it will measure a construct in physical education that is often disregarded and neglected in an era where assessment can be a crucial necessity for program retention and recognition.

CHAPTER 2

LITERATURE REVIEW

Introduction

Assessment is the collection of evidence of student learning that is used to make educational decisions regarding student progress (Lund & Tannehill, 2005). Assessment is a "powerful force in the instructional process" (Lund & Shanklin, 2011, p. 218) and constitutes a variety of purposes and functions. Generally, assessment can be broken into two types: assessment that is used in a formative sense to facilitate student learning, and assessment for accountability, where assessment is used for reporting results, grading, or program evaluation (Lund & Tannehill, 2005). When assessment is used in the formative sense to facilitate student learning, the types of assessment used, and the use of assessment results, differ than those used for accountability. Assessment occurs when students are given the opportunity to demonstrate their level of proficiency and understanding over the course of time and in a variety of contexts and settings (Siedentop & Tannehill, 2000). Rooted within the instructional process, assessment should be aligned to the learning objectives (Chen, 2005; Lund & Veal, 2008; van der Mars et al., 2018). Assessment is used to enhance the teaching-learning process by using assessment results to improve instructional and pedagogical practices, and by providing students with the necessary information required to improve their skills (Black & Wiliam, 2010; Emmanouilidou, Derri, Aggelousis, & Vassiliadou, 2012; Frapwell, 2010; Hay, 2006;

Hay & Penney, 2009; Lund, 1992; Matanin & Tannehill, 1994; Melograno, 1997; Ni Chronin & Cosgrave, 2013; Pickup & Price, 2007; Piotrowski, 2000; Rink, 1993; Siedentop & Tannehill, 2000; Veal, 1988; Wiliam, Lee, Harrison, & Black, 2004; Wright & van der Mars, 2004). The use of assessment to facilitate student learning is a recommended best practice in physical education (The Society of Health and Physical Educators [SHAPE America], 2015).

Student assessment is one of the four Essential Components of Physical Education working synergistically with curriculum, appropriate instruction, and policy and environment as the recommended best practices for teaching physical education (SHAPE America, 2015). Identifying assessment as an essential component of effective physical education indicates that student assessments should be aligned with the national and/or state standards, grade level outcomes, and district curriculum (Lund & Kirk, 2010; SHAPE America, 2015; Thurlow, 2002), regardless of the type of assessment (e.g., diagnostic/preassessment, formative, and summative). This collection of student learning evidence should be used to guide pedagogical practices and decision-making to facilitate student learning (SHAPE America, 2015; Siedentop & Tannehill, 2000). Assessment is most effective when it is used systematically, objectively (Veal, 1988), and continuously throughout the learning cycle to ensure assessment opportunities truly represent student learning across time (Tousignant & Siedentop, 1983).

Types of Assessment

Different types of assessment are most effective when utilized at different stages of the learning process. SHAPE America recommends the use of preassessment (also referred to as diagnostic assessment) at the beginning of an instructional unit to determine

the extent to which students are already familiar with content and skills (SHAPE America, 2015). Assessment data from the preassessment is used to create or modify the planned instructional unit such that the content is developmentally appropriate and challenging enough to progress student learning. Throughout instructional units, it is recommended for practitioners to use formative assessment frequently to check for student understanding and their progression toward the acquisition of the skills and knowledge taught within the unit of instruction. Formative assessment provides students with individualized immediate feedback and information to enhance their progress toward learning. For this reason, formative assessment can be used informally (e.g., verbal feedback without formally recording assessment information) or formally (e.g., recording or writing down assessment data; Rink, 2014; van der Mars et al., 2018) and is often not associated with grading (Lund & Tannehill, 2005). Similar to preassessment, formative assessment provides the teacher with the opportunity to modify instructional practices to meet student needs as evidenced by the assessment data collected about student learning progression (Black & Wiliam, 1988; Desrosiers, Genet-Volet, & Godbout, 1997; Safrit, 1986; SHAPE America, 2015; Veal, 1988). Finally, at the end of an instructional unit, the teacher will use a summative assessment to measure overall student learning and achievement in regard to the unit's intended learning objectives and mastery toward state and national standards (Bailey, 2001; SHAPE America, 2015; Veal, 1988; Wiliam, 2011). Summative assessments can also be referred to as 'assessment of learning' (Lund & Tannehill, 2005). Often, summative assessment results are formally recorded, used for grading purposes, curriculum evaluation, or can be utilized to report to the district or state for accountability purposes, if required to do so (Rink, 2014).

Assessment for Accountability

Apart from using assessment to promote learning during an instructional unit (e.g., diagnostic and formative), assessment can also be used to determine overall success and achievement of instructional programming (Rink et al., 2007; van der Mars et al., 2018). Assessment accountability holds teachers responsible for reporting formal (typically summative) student assessment results to the district or state, thus providing evidence that students are learning and that teaching is effective (Doolittle, 1996; Hay & Penney, 2009; Lund & Kirk, 2010; Lund & Tannehill, 2005; Mintah, 2003; Rink, 2014; Rink & Mitchell, 2002; Safrit, 1986; Veal, 1988). Educators are recommended to use assessment data to track students across grade levels to determine the extent to which students are meeting state and national standards for physical education (SHAPE America, 2015). Additionally, the compilation of assessment data is used to support physical education programming by providing evidence of the quality and effectiveness of the physical education program (Veal, 1988). Historically, physical education has not been held accountable to produce assessment data, and therefore, physical educators are not accustomed or prepared to do so (Lund & Tannehill, 2005; van der Mars et al., 2018). Currently in education in the United States of America, the enactment of the Every Student Succeeds Act (E.S.S.A.) in 2015, provided large impetus for assessment accountability of student learning across many subject areas, including physical education. This legislation considers physical education an important subject contributing to a students' "well-rounded education" (SHAPE America, 2016). Therefore, E.S.S.A. can act as a catalyst for physical education teachers to produce assessment accountability data (SHAPE America, 2016; van der Mars et al., 2018).

Accreditation

Similar to accountability policies at the district and state level, teacher education programs are held accountable for effective programming. To assure the quality and effectiveness of teacher education programming in higher education, most programs opt to become accredited by an agency that is recognized by the U.S. Department of Education (U.S. Department of Education, 2019). The accreditation of institutions and programs in higher education serve the purpose of ensuring that programs in higher education meet acceptable levels of criteria and standards for quality programming, as defined by each accreditation agency (Database of Accredited Postsecondary Institutions and Programs [DAPIP], n.d.). Accreditation agencies, including federally, nationally, non-governmental, and state recognized accreditation agencies, ultimately grant the university permission to provide financial aid to students. Various accreditation agencies utilize different criteria and standards for effective programming (DAPIP, n.d.). Often, accreditation standards address teacher candidates' professionalism, instructional planning ability, assessment practices, and/or community involvement.

Accreditation agencies act as a gatekeeper within an institution or teacher education program to produce effective and qualified future educators. SHAPE America endorses the Council for the Accreditation of Educator Preparation (CAEP) accreditation agency for physical education teacher education programs (SHAPE America, 2017). Accredited physical education teacher education programs are expected to produce eligible candidates for certification in teaching physical education (NCATE, 2008). During CAEP accreditation, the program's curriculum is evaluated and approved using criteria and standards outlined by the Special Professional Association (SPA). Of the 6

standards, The National Standards for Initial Physical Education Teacher Education (SHAPE America, 2017) Standard 5 addresses 'Assessment of Student Learning' by explaining that "physical education candidates select and implement appropriate assessments to monitor students' progress and guide decision making related to instruction and learning." (SHAPE America, 2017, p. 6). Specifically, physical education teacher candidates should be able to utilize authentic formal assessments, formative, and summative assessments, and use reflection of the student assessment results to guide decision-making for the facilitation of student achievement of the intended objectives and standards (SHAPE America, 2017). The extent to which other accreditation agencies measure the effectiveness of teacher education candidate assessment practices is unknown.

Assessment: A Disregarded Practice in Physical Education

Despite assessment serving as a fundamental and vital skill for educators, assessment continues to be problematic pedagogical skill in physical education (Castelli & van der Mars, 2018; Lopez-Pastor et al., 2013; van der Mars et al., 2018) and is even referred to as a "heated discourse among physical education teachers" (Leirhaug & MacPhail, 2015, p. 32). Assessment in physical education is a pertinent but often disregarded topic not only in the research literature (Emmanouilidou, Derri, Aggelousis, & Vassiliadou, 2012; Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013; Redelius & Hay, 2009) but also in practice (Castelli & van der Mars, 2018; Leirhaug & MacPhail, 2015; Lund & Veal, 2008; Kneer, 1986; MacPhail & Murphy, 2017; Park, 2017; Sofo, Ocansey, Nabie, & Asola, 2013; Svennberg, Meckbach, & Redelius, 2018; Tolgfors, 2018; van der Mars, Timken, & McNamee, 2018; Veal, 1988).

Assessment abandonment is most notable at the elementary physical education level (Imwold, Rider, & Johnson, 1982; Ni Chronin & Cosgrave, 2013).

Physical Education Teacher Perceptions of Assessment

Physical education teachers' disdain for assessment practices was primarily identified around the 1970's (Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013), likely due to the lack of national, state, and district accountability in physical education (James, Griffin, & Dodds, 2009; Lund & Tannehill, 2005; van der Mars et al., 2018). Physical education teachers perceive assessment to be one of their most difficult responsibilities due to their lack of knowledge about and competence for using assessment (Kneer, 1986; Lander et al., 2016; Lander et al., 2017; Morgan & Hansen, 2007; Leirhaug & MacPhail, 2015; Matanin & Tannehill, 1994; Veal, 1988), leaving this a skill often omitted entirely (Lander et al., 2016; Lander et al., 2017; Leirhaug & MacPhail, 2015). Assessment is cited as a practice that needs further development for in-service teachers (Borghouts, Slingerland, & Haerens, 2017; Imwold, Rider, & Johnson, 1982; Kneer, 1986; Leirhaug & MacPhail, 2015; Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013; Sofo, Ocansey, Nabie, & Asola, 2013) and preservice teachers (Lund & Veal, 2008). The need for formal (documented) formative assessment is critical to improving the teaching-learning process in physical education (van der Mars, Timken, & McNamee, 2018).

> *"Without such evaluation physical education programs are like travelers with no destination in mind and no compass to guide their journey. They have little or no idea where they are going and won't know where they are once they arrive."* (Imwold, Rider, & Johnson, 1982, p. 17).

Perceived Barriers of Assessment in Physical Education

Overall, the focus of assessment research in physical education has centered less on the frequency in which physical education teachers use assessment and more on the barriers physical education teachers face with respect to using assessment (Morgan & Hansen, 2008; Penney, 2012; Stiggins, 1997). A teacher's assessment practices are likely influenced by his or her perception of assessment processes (Lander et al., 2016; Leirhaug & MacPhail, 2015). Physical education teachers often perceive that assessments are too time consuming, require large amounts of equipment and set-up, involve too many students with too few assessment administrators, take place in inauthentic settings, and require unattainable equipment and facilities (DinanThompson & Penney, 2015; Georgakis, Wilson, & Evans, 2015; Kneer, 1986; Lander et al., 2015; Lander et al., 2016; Lander et al., 2017; Leirhaug & MacPhail, 2015; MacPhail & Halbert, 2010; Michael et al., 2016; Mintah, 2003; Morgan & Hansen, 2007; Veal, 1988; Wiart & Darrah, 2001). With a lack of accountability, district administrative policies, curriculum, and administrator support, physical education teachers often forgo the assessment process altogether (Lander et al., 2015; Michael et al., 2016; Mintah, 2003; van der Mars et al., 2018; Veal, 1988). Additionally, physical education teachers feel they have too many other responsibilities (e.g., duties) aside from teaching physical education, which also present barriers to performing assessment (Mintah, 2003). Most notable barriers include a physical education teacher's poor conception and understanding of assessment and its purpose and benefit in education, as well as a lack of assessment training, which leaves teachers underprepared to perform assessment tasks (Annerstedt & Larsson, 2010; DinanThompson & Penney, 2015; Imwold, Rider & Johnson, 1982; Lander et al., 2015,

2016, 2017; Lund & Veal, 2008; Michael et al., 2016; Slingerland et al., 2017; van der Mars et al., 2018). Furthermore, a study with physical education teachers in California found that teacher buy-in and philosophy were barriers to the teachers use of standards-based assessment (Michael et al., 2016). Physical education teachers desire efficient, simple, and authentic assessments that can be used easily in a natural physical education setting in conjunction with the necessary professional development training and support to become successful assessors (Chen, 2005; Haynes & Miller, 2015; Imwold et al., 1982; James et al., 2009; Landers et. al., 2016; MacPhail & Halbert, 2010; Michael et al., 2016; Mintah, 2003; Ni Chroinin, & Cosgrave, 2013; Oh, Graber, & Woods, 2016; Patton & Griffin, 2008; Richard, Godbout, Tousignant, & Grehaigne, 1999; Rink, Jones, Kirby, Mitchell, & Doutis, 2007). Regarding the use of standards-based assessment in the United States of America, only one study investigated teacher perceptions of assessing student learning while utilizing standards-based assessment practices (Michael et al., 2016). Thus, there has been limited attention or concern for standards-based assessment in physical education. Due to low perceptions of assessment in physical education, assessment is often left out of the teaching-learning process (Lander et al., 2016; Lander et al., 2017; Leirhaug & MacPhail, 2015).

Frequency of Assessment Use in Physical Education

The lack of concern for assessment in physical education is alarming (Imwold, Rider, & Johnson, 1982; Kneer, 1986). Assessment, if used at all, is often used inappropriately (Borghouts, Slingerland, & Haerens, 2017). Examples of misuse include utilizing only summative assessments, assessing students on non-instructional tasks (e.g., attendance and dressing out), assessing using subjective criteria rather than skill testing

with specific criteria, and the failure to align assessment purposes with student learning outcomes (Borghouts, Slingerland, & Haerens, 2017; Hay & Macdonald, 2008; Imwold, Rider, & Johnson, 1982; Lund & Veal, 2008; van der Mars et al., 2018).

Assessment takes place most frequently at the junior high level and least frequently at the elementary level, occurs more often in larger districts as compared to smaller ones (Imwold, Rider, & Johnson, 1982), and is conducted more frequently by female physical education teachers than by males (Imwold, Rider, & Johnson, 1982; Kneer 1986; Mintah, 2003). Teachers with more years of teaching experience tend to assess more frequently than beginning teachers (Desrosiers, Genet-Volet, & Godbout, 1997; Imwold, Rider, & Johnson, 1982; Kneer, 1986), and teachers who coach assess more frequently than non-coaches (Imwold, Rider, & Johnson, 1982). Performance-referenced assessment are most commonly used, where teachers compare students against their peers, rather than using criterion-referenced assessment (using criteria to evaluate), except at the high school level where criterion-referenced assessment was more widely used (Imwold, Rider, & Johnson, 1982). Most often, assessments were comprised of subjective criteria to measure attendance, behavior, or dressing out for physical education (Borghouts et al., 2017; James, Griffin, & Dodds, 2009; Imwold, Rider, & Johnson, 1982; Matanin & Tannehill, 1994; Morrow, 1978; Redelius & Hay, 2012; Sofo et al., 2013). When measured, elementary physical education teachers were the least likely to assess, especially for the use of skills tests (Imwold, Rider, & Johnson, 1982) but conversely, were found to use authentic assessments more frequently than other secondary teachers (Mintah, 2003). Possible reasons for elementary teachers assessing less frequently overall could be the large class sizes with minimal physical education

class times (i.e., only a half hour; Imwold, Rider, & Johnson, 1982). Regarding resources for teachers to reference for assessment purposes, there were few elementary physical education textbooks and curriculums that addressed objective assessments that are appropriate for elementary school students (Imwold, Rider, & Johnson, 1982). At all school levels, informal assessment practices, such as teacher observation (Borghouts et al., 2017; Mintah, 2003; Sofo et al., 2013), and informal feedback were the most commonly used assessment procedures (Lander et al., 2015) with formal assessment used least frequently (Borghouts et al., 2017; Veal, 1988). In regard to the use of recommended assessment practices, the use of diagnostic, formative, and summative assessments (Desrosiers et al., 1997; Veal 1988), skill testing (Desrosiers et al., 1997; Imwold, Rider, & Johnson, 1982; Kneer, 1986), and formal assessment did take place in some physical education settings (Sofo et al., 2013).

Consistent with studies of in-service physical education teachers, studies that examined preservice and student teachers' assessment practices found that preservice physical education teachers described similar barriers and demonstrated a lack of appropriate assessment practices (Goc Karp & Woods, 2008; Lund & Veal, 2008; Morrow, 1978; Rink et al., 1994). A notable finding was a lack of alignment between assessment and instructional goals, thereby providing evidence of needed intervention in teacher education programming (Borghouts, et al., 2017; Lund & Veal, 2008; van der Mars et al., 2018; Veal, 1988).

Current Frequency of Assessment Literature is Not Enough

Research regarding the frequency and types of assessment used in physical education have most frequently been conducted outside of the United States of America,

with studies conducted in Canada (Desrosiers, Genet-Volet, & Godbout, 1997), England (Carre et al., 1983), Spain (Arias-Estero & Castejon, 2014), South Korea (Park, 2017), Ireland (Murphy & O'Leary), Norway (Leirhaug & MacPhail, 2015; Leirhaug, MacPhail, & Annerstedt, 2016), Netherlands (Borghouts, Slingerland, & Haerens, 2017), Australia (Landers et al., 2015, 2016), and Africa (Sofo, Ocansey, Nabie, & Asola, 2013). Although these investigations provide great overall context regarding assessment in physical education, increased research is needed to understand physical education teacher assessment in the United States context. For instance, whereas several of the above-mentioned countries have national curriculums to guide assessment practices, policies, standards, criteria, and teaching philosophies, the United States of America does not; rather, educational decisions are made at the state and district level. Therefore, the ability to generalize the findings from the assessment frequency studies to physical education in the United States of America is limited.

Within the United States of America, only two frequency studies (van der Mars et al., 2018a, 2018b) have been conducted since the national physical education standards were enacted in 1995 by the National Association for Sport and Physical Education (NASPE; currently SHAPE America). Hence, the political climate and context for assessment best practices in earlier studies (see Hensley, Lambert, Baumgartner, & Stillwell, 1987; Imwold, Rider, & Johnson, 1982; Kneer, 1986) differ greatly from studies that took place after the enactment of the national standards in 1995 and the recommendation for use of standards-based assessment (SHAPE America, 2015). However, regardless of the political climate and enactment of a country's standards and curriculum, physical fitness is consistently the most widely assessed content in physical

education across the world (Lopez-Pastor et al., 2013) leaving out all other important topics and content areas in physical education. Consequently, ineffective assessment practices persist, despite the recommendation for the use of standards-based assessment, best practices regarding student assessment with the Essential Components of Physical Education, and the accreditation requirements for teacher candidates to effectively assess student learning (SHAPE America, 2017).

Assessment Recommendations for Preservice and Inservice Teachers

Since the 1980's, emphasis on preservice teacher training to improve assessment practices has been heavily recommended. Experts recommend the creation of more opportunities for preservice teachers to practice using assessment in more authentic settings that are better representative of real-world classrooms (i.e., with large class sizes, little time, and few resources; Imwold, Rider, & Johnson, 1982). Recommendations to improve assessment practices also include increased professional learning for in-service teachers (Chen, 2005; Lander et al., 2015; Lander et al., 2016; Imwold, Rider, & Johnson, 1982) and the development of efficient, simple, and large-group type assessments for physical education teachers to use (Imwold, Rider, & Johnson, 1982). Training physical education teachers to use assessment has been successful at mitigating poor assessment practices in the past (Kneer, 1986; Lander et al., 2015, 2016, 2017; Michael et al., 2016).

Assessing SHAPE America Standard 1: Motor Skills

As previously noted, fitness testing is currently the most prevalent assessment practice taking place in physical education (Lopez-Pastor et al., 2013), thereby presenting shortfalls in the assessment of more diverse content in physical education, particularly at the elementary level (Imwold et al., 1982; Ni Chronin & Cosgrave, 2013). SHAPE

America's National Standard 1 states "The physically literate individual demonstrates competency in a variety of motor skills and movement patterns." (SHAPE America, 2013). Motor skills are the most fundamental aspects of the psychomotor domain (Harrow, 1972; Simpson, 1972) in physical education comprised of locomotor skills (e.g., run and jump), non-locomotor skills (e.g., balance and weight transfer), and manipulative skills (e.g., throw and catch; SHAPE America, 2013). Motor skill competency will be the focus of this paper as motor skill development at the elementary school level is pertinent to more complex skill development (Logan et al., 2015; Lubans et al., 2010) showing a positive and bidirectional relationship between motor skill competence, physical activity, and health-related fitness levels as an individual grows older (Lima et al., 2017a, 2017b).

Constituting major importance in elementary physical education, motor skills are a primary indicator within the SHAPE America's Grade-Level Outcomes (SHAPE America, 2013). The grade-level outcomes break down each of the national standards into performance indicators that are achievable by grade level. In reference to this document, motor skills should be mastered by the completion of the fifth grade thereby acting as a catalyst for the measurement of student levels of performance of motor skills (SHAPE, 2013). However, there are currently no studies conducted in the United States of America regarding physical education teachers' practices or perceptions of assessing students' motor skills at the elementary level. Therefore, this dissertation will focus primarily on the assessment practices and perceptions of elementary physical education teachers regarding the assessment of student achievement toward the National Standard 1: Motor skills, and/or the aligned motor skill standard at the state level.

Need for Valid and Reliable Survey Tool

In order to gain a wide understanding of assessment practices and perceptions of motor skill assessment from teachers across the United States of America, we propose that a quantitative survey to measure the motor skill assessment behavior construct is most ideal. A thorough search of the literature has demonstrated that there currently are no existing surveys that have been used specific enough for the purpose of this study. Most of the studies measuring practices and perceptions of assessment used qualitative methods to interview or observe physical education teachers with few using quantitative measures, including surveys. To most appropriately understand survey results, surveys used for quantitative measures should produce valid and reliable measurements (i.e., psychometric properties) for the specific population that it is studied (Bandalos, 2018). Although many of the previously mentioned studies examined the reliability and validity of the measures within their populations, the context and populations for those studies differ from those targeted for this study (i.e., United States-based elementary physical education teachers). Therefore, many of the variables previously measured, survey items, and themes (both perception and frequency) within the literature have informed the variables examined in this dissertation, which was geared toward elementary physical education teachers in the United States of America regarding their practices and perceptions of motor skill assessment.

The Purpose of This Research

To measure current elementary physical education teacher assessment practices and perceptions of assessment in an exploratory way, we developed a survey with content validity and internal consistency reliability using the major frequency and perception

themes and variables from the aforementioned physical education assessment literature. The major themes from the literature informed the development of survey items within the assessment behavior survey.

The purpose of this dissertation was threefold. First, we developed survey items that validly measure elementary physical education teachers' motor skill assessment behavior (Study 1). Next, we administered the perception of motor skill assessment survey subscale nationwide to elementary physical education teachers to obtain psychometric (internal consistency reliability and factor structure) properties (Study 2). Finally, we used the data from the survey tool to describe teacher perceptions of motor skill assessment and make predictions for the improvement of motor skill assessment behavior (Study 3).

The research questions that drove this dissertation include:

RQ1. What is the content validity of the Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey as informed by a panel of content experts and inservice elementary physical education teachers?

RQ2. What are the psychometric properties of the Elementary Physical Education Teacher Perception of Motor Skill Assessment Survey subscale, based on data collected from a sample of elementary physical education teachers across the United States of America?

RQ3a. To what extent do assessment perceptions differ based upon whether teachers assess motor skills or not?

RQ3b. Which motor skill assessment perception factors predict the likelihood for assessment of motor skills?

CHAPTER 3

STUDY 1: SURVEY DEVELOPMENT AND CONTENT VALIDATION USING THE MODIFIED DELPHI METHOD FOR THE ELEMENTARY PHYSICAL EDUCATION MOTOR SKILL ASSESSMENT BEHAVIOR SURVEY

Introduction

The Society of Health and Physical Educators America (SHAPE America) includes assessment as one of the four Essential Components of Physical Education (SHAPE America, 2015). Quality physical education teachers should collect evidence of student learning toward meeting national and/or state level standards and grade-level outcomes (SHAPE America, 2015). However, assessment as an instructional practice is often a forgone behavior in physical education (Lander et al., 2016, 2017; Leirhaug & MacPhail, 2015; Lund & Veal, 2008).

Assessment Accountability

Unlike other academic content areas, state and national accountability systems have not consistently held the physical education profession accountable for producing evidence of student learning assessment data (van der Mars et al., 2018), thus leaving the assessment behaviors of teachers unmonitored. Theoretically, with assessment as an unmonitored process, a teacher's assessment behavior is likely influenced by his or her

perception of the behavior (e.g., ease of administration, benefits for students; Lander et al., 2015, 2016; Leirhaug & MacPhail, 2015).

Perceptions of Assessment

Physical education teachers typically perceive assessment as an arduous and time-consuming task (DinanThompson & Penney, 2015; Georgakis, Wilson, & Evans, 2015; Lander et al., 2016, 2017; Leirhaug & MacPhail, 2015; MacPhail & Halbert, 2010; Michael et al., 2016), and often have misconceptions about the purpose and function of assessment (Kneer, 1986; Matanin & Tannehill, 1994; Veal, 1988). There is much variability within the methodologies used in the extant literature which places implications for generalizability at a disadvantage. Methodologies within the current perceptions of assessment in physical education literature span qualitative (n = 7), quantitative (n = 2), and mixed methods approaches (n = 6). Of these research studies, most were focused on teacher perceptions of a specific type of assessment tool or were specifically designed to measure assessment perceptions after some form of intervention (Fisher et al., in preparation). Notably, the majority of the perception studies took place outside the United States of America, leaving the contextual understanding of standards-based assessment amiss (Fisher et al., in preparation).

Frequency of Assessment

There is also a large gap in the literature surrounding physical education teacher use (frequency) of assessments in physical education. To our knowledge, there are recently only four studies in the United States of America that measured the extent to which physical education teachers are using assessment. Michael et al. (2016) surveyed 309 middle school physical education teachers' use of standard-based assessment in

California. van der Mars and colleagues (2018) developed and tested a systematic observation tool to measure physical education teacher formal assessment practices. Lund and Veal (2008) measured preservice teacher assessment practices during a student teaching semester. Similar to other frequency articles, one from the 1980's and one from Ireland, the most notable findings across the frequency studies was the reduced amount of assessment practices taking place at the elementary level (Imwold, Rider, & Johnson, 1982; Ni Chronin & Cosgrave, 2013). Compounding the generalizability across studies, there was large variation in methodologies used, types of physical education teachers measured (preservice and inservice), and differing research questions and purposes. Furthermore, studies used relatively small sample sizes and were conducted within confined geographic locations. Consequently, there is a large gap in the research on physical education teachers' use of assessment practices, particularly standards-based assessment.

Motor skills are the major component within the SHAPE America National Physical Education Content Standard one, stating "The physically literate individual demonstrates competency in a variety of motor skills and movement patterns" (SHAPE America, 2013). Within the SHAPE America grade-level outcomes, motor skills are a focus for instruction and assessment from Kindergarten to 5th grade (SHAPE America, 2013). With the national standards and grade-level outcomes informing physical education teacher practices across the nation, elementary physical education is an ideal setting to teach and master the fundamental motor skills (Lander et al., 2015; e.g., jump, skip, catch), which provide the foundation for more complex movement (Logan et al., 2015). Contrary to the typical 'leave assessment out' of physical education culture (Lund

& Veal, 2008), the SHAPE America Research Council supports an initiative targeting the assessment of student motor skill competence across the K-12 setting (Castelli & van der Mars, 2018). This initiative provides impetus to begin measuring teacher's assessment practices relating to motor skills.

The extant research on physical education teachers' assessment practices and perceptions lacks studies that focus on the assessment of motor skills at the elementary level in the United States. Therefore, this study will focus on SHAPE America's National Standard 1, motor skills, and the assessment of motor skills at the elementary level. Using the perception and frequency themes found within the physical education assessment literature (Fisher et al., in preparation) as a guide, the purpose of this study was to develop and explore the content validity of new survey items surrounding physical education teacher assessment behavior (e.g., frequency and perceptions) of motor skills that can be used with elementary physical education teachers across the United States of America. The research question for this study was, "What is the content validity of the Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey as informed by a panel of content experts and current quality inservice elementary physical education teachers?"

Methods

Design

This study used a sequential, quantitative, descriptive-analytic, modified Delphi Method research design (Avella, 2016) using open- and close-ended survey questions. The modified Delphi method is a commonly used research methodology that utilizes several rounds (Hsu & Sandford, 2007) to obtain anonymous consensus from content

experts surrounding a topic of contemplation or to define an ambiguous term or idea (Avella, 2016). In comparison to a traditional Delphi method, the modified Delphi can utilize literature reviews as the theoretical underpinning to the content under scrutiny (Avella, 2016; Bandalos, 2018). For the purposes of this project, the topic of contemplation surrounded the creation of appropriate survey items to help understand the behavioral constructs underlying assessment in physical education (i.e., frequency of motor skill assessment use and perceptions of motor skill assessment). We derived survey items from the emerging themes of two extensive literature searches surrounding teacher frequency and perceptions of assessment in physical education and their recommendations for improving assessment behavior (Fisher et al., in preparation). Using the modified Delphi method, we presented experts with survey items and asked them to provide anonymous feedback (Rowe & Wright, 1999) in the form of rating items on relevance (accuracy of measuring construct) and clarity (understandability) and making recommendations for ways to improve survey items. After systematically compiling ratings and feedback from the experts, the research team made modifications to the item wording, phrasing, or response option, and send the items back out to experts for another round. Sometimes the ratings and feedback deemed item removal or the addition of new items appropriate. Typically, to avoid participant attrition and exhaustion (Schmidt, 1997), the Delphi method lasts for a maximum of three rounds, typically using at least two rounds or until consensus is reached (Keeney et al., 2005). Consensus ratings can range from 55% - 100% total agreement among experts, but commonly 70% (Vernon, 2009) or 75% (Diamond et al., 2014) is an appropriate standard. To strive for rigor, the

research team declared 80% total agreement as the appropriate measure for consensus in this study to obtain item content validity (Linstone & Turoff, 2002).

Sample

Participants (N = 25) included a purposive sample of content experts in the field of physical education assessment (n = 14; Table 3.1) and current inservice elementary physical education teachers (n = 11; Table 3.2).

Sample Selection

We selected content experts based on several factors, including (a) international and national publication and presentation history in the field of assessment in physical education, (b) geographical constraints, such that individuals know and understand the cultural aspects related to assessment in the United States of America, and (c) the expert's vested interest in improving the field of assessment in physical education. We purposefully selected inservice elementary physical education teachers based on (a) their knowledge of the importance of assessment in physical education, (b) their current use of assessment in authentic physical education settings, (c) the grade level for which they teach (e.g., elementary), and (d) their overall standing with the university as a student teaching mentor. We evaluated each expert and inservice teacher for expertise using the 'Expert Rater Table' (see Table 3.1 & 1.2) to obtain descriptive information for the modified Delphi participants.

Instrumentation

We created all initial survey items from two systematic reviews of the literature about teacher frequency and perceptions of assessment in physical education (Fisher et al., in preparation). The emerging themes and variables underlying teacher frequency and

perceptions served as the item themes during survey development (Avella, 2016; Bandalos, 2018). Themes from the systematic reviews that were included in this survey surrounded frequency of assessment, types of assessment used, when assessment is used, resources used, learning opportunities, and perceived facilitators and barriers for using assessment that would elicit assessment behaviors (i.e., frequency of use).

Frequency and Perception Survey for Delphi

The initial survey consisted of 50 items surrounding the topics of assessment behavior (n = 1), perceived assessment importance beliefs (n = 2), assessment frequency (n = 20), likelihood of conducting assessment (n = 3), perception of assessment (n = 17), learning opportunities for assessment (n = 4), and three Every Student Succeeds Act (E.S.S.A.) items. Each item, aside from the E.S.S.A. items, has a focus on motor skills. The demographics and school context questions were not included in the modified Delphi process but rather were created and evaluated by the research team. The assessment behavior item asked, 'Do you assess the motor skills included in the SHAPE America Standard 1: "demonstrates competency in a variety of motor skills" (i.e., locomotor [gallop] and object control [throw]; SHAPE America, 2013) in physical education?' with a 'yes' or 'no' option to direct participants to different parts of the survey based on their response. Perceived assessment importance belief item stem asks 'To what extent do you agree with the following statement…' with item examples including 'Motor skills (i.e., skipping and throwing) should be taught in elementary school physical education' and 'Motor skills should be formally (written/documented) assessed in elementary physical education (i.e., motor skill performance is recorded/documented for each student)' on a six-point Likert-type response scale (1 = Strongly Disagree, 3 = Somewhat Disagree, 4 =

Somewhat Agree, 6 = Strongly Agree). Item examples referring to frequency of assessment practices include question stems such as “To what extent do you use the following types of motor skill assessment, where ‘always’ represents ‘during every lesson’”, with item examples including ‘teacher observation (silently observing)’ or ‘teacher observation with verbal/non-verbal feedback (or prompts) to students’ (pseudo-assessment, Lund, 1992). Frequency items used a six-point Likert-type response scale (1 = Never, 3 = Rarely, 4 = Occasionally, 6 = Always). The likelihood items stem asked “To what extent do you agree with the following statements…I would be more likely to assess…” with an item example ‘if the district held me accountable to assess’ on a six-point Likert scale (1 = Strongly Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 6 = Strongly Agree). Item examples referring to perceptions of assessment include questions such as “Relative to using formal assessments of motor skills, to what extent do you agree with the following statements?”, with items including ‘My class sizes are too large to assess’ and ‘I have sufficient time to assess motor skills’. Perception items used a six-point Likert-type response scale (1 = Strongly Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 6 = Strongly Agree). To glean the degree to which teachers perceive learning opportunities as beneficial to learning about assessment, the item stem for learning opportunities stated “To what extent do you agree that the following learning opportunities would make you more likely to use motor skill assessment?”. Example learning opportunity items include ‘physical education conferences’ and ‘professional development or inservice training for physical education assessment’ with responses on a six-point Likert scale (1 = Strongly Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 6 = Strongly Agree). Finally, we wanted to understand the extent to which

physical education teachers were informed and aware about E.S.S.A. funding by asking questions like ‘Does your district receive funding from the E.S.S.A. for physical education?’ and ‘Do you know how to advocate for your physical education program to receive E.S.S.A. funding?’ on a ‘yes, no, and I do not know’ response scale. The survey also includes demographic and school context questions, including teaching responsibilities, school social economic status, state employed, Title I school, etc. The research team created the demographic and school context items and therefore, did not include these items on the modified Delphi survey to reduce the number of items participants would need to rate.

Delphi Survey

Consistent with the nature of a modified Delphi method, the expert raters read the initial survey items and rated them on a four-point Likert Scale for relevance and clarity (i.e., 1 = Very Irrelevant, 2 = Irrelevant, 3 = Relevant, 4 = Very Relevant; 1 = Very Unclear, 2 = Unclear, 3 = Clear, 4 = Very Clear) Experts provided feedback and commentary for any recommended change under the comments section. See Figure 1 for an example of the modified Delphi rating scale for an importance item.

Procedures

We acquired Institutional Review Board approval from the University of South Carolina prior to the start of this study. We recruited a purposive sample of experts (content and inservice teachers; Avella, 2016; Rowe & Wright, 2001) using their work email in order to elicit individuals with the best content knowledge and experience regarding assessment in the field of physical education. Initially, we emailed an invitation to participate describing the entire project and the expert’s potential role in the project.

Upon consent, experts completed the 'Expert Rater Table' (Table 3.1 & 1.2). Any expert who declined to participate was removed from the pool and a new expert was considered by the researchers. After we reached a desired sample of more than 10 content experts and 10 inservice teachers, we began the modified Delphi process. The modified Delphi took place in two phases, one with content experts and the second with inservice elementary physical education teacher experts. The same process of recruitment took place for both populations, however the inservice experts did not receive the survey until after it was revised and finalized through consensus with the content experts (described below).

Content Expert Phase

Content experts in the field of physical education assessment only received the survey items upon completion of the Expert Rater Table (Table 3.1). Content experts received the survey items first, before the inservice teachers. During each round of the modified Delphi process, each expert was asked to rate every survey item on relevance (i.e., how important the item is to measuring the assessment behavior constructs) and clarity (i.e., ability to understand what the item is asking) using a four-point Likert-type scale (e.g., 1 = Very Irrelevant, 4 = Very Relevant; American Education Research Association, 2014). All expert ratings and comments were systematically documented and collated for each round. Any item with a mean average of 3.2 (80%) or below for any given category (relevance or clarity) was revised using the expert feedback. The revised items, controlled feedback, and consensus average by item from that round were redistributed to experts for a subsequent round of rating and feedback. Experts required

two rounds to obtain consensus. Experts were reminded via email on a two-week basis to complete each round and/or sent a ‘thank you’ for participation.

Inservice Physical Education Teacher Phase

Inservice physical education teachers received the Delphi survey after the content experts obtained consensus with a goal to ensure interpretability, utility, and authenticity in the field of elementary physical education. The inservice teachers did not receive the survey items until completion of the Expert Rater Table (Table 3.2). The modified Delphi rating and revision process for survey items was the same for the inservice teachers as it was for the content experts. The inservice teachers participated in two rounds to obtain consensus. Teachers were reminded via email on a 2-week basis to complete each round or received a “thank you” for participation.

Content Validity

Content validity measures the degree to which the items represent the latent variable to be measured (Benson & Clark, 1982). We revised survey items using content expert and inservice teacher relevancy and clarity ratings and feedback in order to finalize the survey, thus providing us with content validity (Linstone & Turoff, 2002). Davis (1992) recommends content validity of 80% or higher when using a panel of Delphi experts. Consensus between raters for each item with a mean rating at or above 3.2 on the four-point Likert scale (80%) demonstrated content validity of survey items.

Data Analysis

We calculated means and standard deviations for each item, by each rater, and as a grand mean for each theme (frequency and perception) and scale (relevance and clarity) separately for content experts and inservice teachers. We decided that acceptable means

(i.e., consensus) by item would need to be 3.2 of the four-point Likert scale (80%) or higher by individual item or else the item was flagged for revision. Any individual rating below 3 (e.g., 2 or 1) was also flagged and revisited, using expert and teacher comments, to gain greater understanding about how to improve the item.

Delphi participants had the option to leave a comment for any item that they recommended for revision. We systematically recorded every comment that was made by an expert or teacher for each item. We used Delphi participant comments to revise any item that received a consensus score below the standard (3.2 average by item, or 80%) and to gain greater understanding of the Delphi participant perspective.

Results

Content experts reached consensus after two rounds of the modified Delphi process and were provided two "final approval" rounds. The first final approval occurred after the content experts reached consensus and we made the recommended changes on Round Two using the modified Delphi process. The experts were given a second chance for final approval after the inservice teachers made recommendations to the survey and before it was to be entered into its final online format. After the content expert's first chance at final approval, the survey was revised based on their comments, and sent to inservice elementary physical education teachers for Delphi ratings. The inservice teachers reached consensus using only one round of Delphi ratings. The survey was further revised based on inservice teacher recommendations for relevance and clarity. After inservice teachers made recommendations, the research team made revisions to item type and type of data gleaned from the survey to ensure statistical utility (i.e., reliability). Finally, we sent the finalized survey to our content experts for their final

approval. This final approval from content experts served as a last chance to change items prior to entering the items into its online format for dissemination.

Content Expert Criteria

Table 1 outlines the descriptive data related to content expert criteria for participation. The majority of content experts (*n* = 11) had taken at least one university-wide course about assessment during their university training. Forty-three percent (*n* = 6) reported having taught a university course on assessment and 43% (*n* = 6) reported that they currently teach and/or published textbooks about assessment in physical education. Most participants (*n* = 11) indicated they consistently use a variety of formal and informal assessments, including diagnostic, formative, and summative assessments. All but one content expert reported having published in either a national or international top-tier research journal about assessment (*n* = 9), national or international lower-tier journals (*n* = 5), and/or a practitioner-oriented journal (*n* = 7). Sixty four percent (*n* = 9) of the content experts were part of a state-wide assessment project in their respective states. Every participant possessed a doctoral degree, with 10 working in university or college settings, one employed by a school district, and three retired with professor emeritus titles. Of the 10 university or college professors, seven were full professors.

Round 1 Content Experts

We sent content experts the initial survey with 50 items relating to assessment behavior (*n* = 1), perceived importance (*n* = 2), frequency (*n* = 20), likelihood (*n* = 3), perception (*n* = 17), learning opportunities (*n* = 4), and three E.S.S.A items (See Table 3). We asked the experts to rate survey item relevance and clarity and provide feedback

for improvement on either relevance or clarity. Finally, consistent with the modified Delphi method, experts could recommend adding or removing any item.

Relevance

Due to low consensus (below 3.2 average or 80% agreement), perception item 33 (M = 3.14, SD = 1.10; e.g., "Relative to using formal assessments of motor skills, to what extent do you agree with the following statements: the set up for assessment is practical" on six-point Likert Scale, strongly disagree to strongly agree) needed revision for relevance. Specifically, the item did not appear relevant because several terms (e.g., set up and practical) needed clarifying and describing.

Clarity

Overall, content expert raters flagged more items for clarity than they did relevance and made recommendations on ways to make the items more understandable and practical for elementary physical education teachers.

Due to low consensus (below 3.2 average or 80% agreement), the assessment behavior item (item 1, M = 2.86, SD = 0.95; yes, I assess; no, I do not assess) needed clarification. Experts recommended revising the item by removing the "SHAPE America standard\" wording and simply referring to "motor skills" to avoid excluding states who create their own standards from the SHAPE America standards. Experts also recommended including a form of assessment "quality" terminology to avoid over- or under-estimating assessment behavior (e.g., does the teacher assess all motor skills or just some? If the teacher claims they do assess, how do you tell if it is quality assessment?).

Three frequency items (items 4, 5, and 9) were flagged for clarity revision. Frequency Item 4: "Which grade levels do you formally (written/documented) assess the

following motor skills? Place an X in the box for each of the grade level bands that you assess those skills. If you do not assess those skills, leave the box blank." The response option was "check all that apply" from Kindergarten to grade six for all locomotor skills, manipulative skills, balance, and weight transfer (from the SHAPE America Grade-Level Outcomes, 2013) (*M* = 2.79, SD = .89). Item 4 needed clarity revision. General recommendations surrounded the types of skills included and formatting of the item to make it more understandable. Frequency Item 5: "Typically, what percentage of class time is spent on motor skill assessment per lesson?" with response options "a) 0 - 10%, b) 11 - 20%, c) 21 - 30%, d) 31 - 40%, e) 41 - 50%, f) More than 50%" (*M* = 2.57, SD = .76) needed clarity revision. The experts recommended clarifying the aspect of 'time' since most teachers do not formally assess daily and the amount of time in physical education class will differ between teachers. Frequency Item 9: "To what extent do you use the following motor skill assessment tools: An assessment tool from adapted physical education" with response options on six-point Likert scale (1 = Never, 6 = Always) (*M* = 3.14, SD = .95) needed clarity revision with experts recommending a 'check all that apply' format instead of a Likert type option.

Experts needed clarification on one likelihood item. Item 25 (*M* = 3.07, *SD* = 1.07) needed revision for clarity. The item asked about the use of curriculum to guide assessment. The experts wanted revision to clarify 'which curriculum' (e.g., district, state, or other) and to consider how the respondent would answer if there was no curriculum in place or available to teachers.

Experts flagged two perception items for clarity revision. Item 33 (*M* = 2.36, *SD* = 1.15) asks about teacher's perceived influence over assessment as impacted by item

wording ‘the set up for assessment is practical’. Experts recommended clarifying both ‘set up’ and the word ‘practical’ asking if we were referring to equipment, space, or both. Item 34 (M = 2.86, SD = .95) was flagged for clarity and asks ‘the amount of equipment needed to assess is practical’ for the same reasons as item 33.

We reviewed expert comments for each item regardless of the consensus rating to glean a greater understanding for the relevance and clarity of each item. The comments also provided evidence of the differing expert perspectives and philosophies regarding the topic of assessment in physical education. Expert comments were used to determine if an item needed to be removed, split into several new items, or needed complete revision.

Reducing and Adding Items

Experts recommended deleting eight items due to low relevancy (via comments, not means and standard deviations; i.e., item is relevant and clear but is addressed within or takes away from another item), redundancy of item content, or because they preferred the item to be split into multiple items. Of those deleted, four were frequency items and three were perception items.

Experts recommended adding 30 items to the survey. One recommendation was to add a ‘teaching motor skill’ question to understand whether teachers even teach the skills, before determining whether they assess the skills in the assessment behavior theme. Two importance items were added to better clarify ‘motor skills’ by breaking the various motor skill components apart (i.e., keeping locomotor, manipulative, and non-locomotor skills separate). Twelve frequency items were added to get a better understanding for the types of assessments and resources teachers are using for motor skill assessments. One likelihood item was added to allow for a qualitative response by asking “other things that

would make me more likely to assess". Nine perception items were added to gather more examples of elements that might act as a facilitator or barrier to assessment. Finally, the experts recommended adding five learning opportunity items to differentiate between state and national conferences, clarifying on continuous professional development versus a 'one-off' version, and other types of materials a teacher might use to learn more about motor skill assessment.

Round 2 Content Experts

After completion of Round 1 revisions, the Round 2 survey consisted of 72 items surrounding assessment of motor skills (behavior; n = 2 items), perceived assessment importance (n = 4), assessment frequency (n = 27), likelihood to assess (n = 4), the perception of assessment of motor skills (n = 23), learning opportunities (n = 9), and three E.S.S.A. items (See Table 3.3). Experts received a revised set of items that explained overall consensus by item for relevance and clarity from the previous round, and an explanation of how each item was revised based on expert comments and recommendations. Any item that was deleted during Round 1 was included (yet crossed out) in this version with consensus rating and explanation for removal. Any item that was added was noted "new item" with an explanation for the addition (i.e., reviewer commentary recommending the addition). The experts were expected to rate relevance, clarity, and provide feedback in the comments section for all included items, just as they had for Round 1. All but one reviewer completed Round 2.

Relevance

Only one frequency item (item 55) was flagged for low relevancy consensus. Item 55 (M = 3.15, SD = .99) asked about the frequency with which teachers used "teacher

observation without a rubric". Experts commented that this is a typical behavior of all teachers, so this item would have a difficult time discriminating between individuals. This item was flagged for removal.

Clarity

Two frequency items (item 4 and 5) were flagged for low consensus rating for clarity. Items 4 and 5 were also flagged for clarity during Round 1. Experts flagged item 4 (M = 2.96, SD = .92) for clarity because this item was extremely overwhelming with all of its categories (i.e., skills listed vertically with checkboxes for each grade level and environment [open vs. closed] and quality [process vs. product] horizontally options for each skill and grade level). The experts recommended simplifying this item by narrowing its focus on skills by grade level and leaving the environment and quality out. In short, experts generally thought that item 4 was 'much too long', too confusing, and that the list of skills included was either too inclusive or not inclusive enough while referring to motor skills (i.e., include non-locomotor skills, include open versus closed environments, etc.). For simplicity, item 4 was changed to its previous version to reflect the simplicity from Round 1 but included a simpler layout to avoid confusion. Frequency item 5 (M = 2.77, SD = .93) remained an item that needed clarification after Round 1. This item continued to need clarification because the item was changed to refer to a 'typical day' in physical education to make the item more relevant to any physical education teacher, regardless of the amount of time they actually get to spend with students. However, 'typical day' remained too ambiguous for most experts, noting that there is no 'typical day' in elementary school since the grade levels that a physical education teacher teaches can change drastically from one day to the next. To be as clear and inclusive as possible,

item 5 was changed to “Within each class that you teach, what percentage of class time do you spend using formal (written/documented) assessment while teaching motor skills?” with several response options ranging from 0% to more than 50%.

Experts flagged perception item 42 (*M* = 3.08, *SD* = .95), which asks if the following statement served as an influence over their assessment behavior; ‘people in my district would notice if I did not assess motor skills (i.e., I am held accountable)’. The experts wanted clarification on who “people” refer to, would like examples, and provided a list of possible ways to rephrase. This item was removed because the ‘accountability’ construct was addressed in other items that did not receive low relevance and clarity scores (e.g., items 22-24).

Similar to Round 1, we reviewed expert comments by item regardless of consensus rating. We made item-level decisions based on the majority of the reviewers’ recommendations, whether they recommended item removal, addition of new items, or revision to item clarity. Aside from the consensus ratings, the qualitative commentary provided considerable detail for item-level decision making.

Reducing and Adding Items

During Round 2, the experts did not recommend removing or adding any items using the consensus ratings, but instead alluded to such in the comments section. At this point, the expert reviewers provided detailed recommendations for combining five items based on redundancy, deleting 13 items for low ratings or because the content was not pertinent to assessment behavior, and removing two items because they were addressed in the demographics section of the survey. The experts also recommended adding a total of 12 items to the survey, including three importance items, two assessment environment

questions (i.e., process versus product and assessment of solo student versus in-groups), five learning opportunity items, one frequency item, and one E.S.S.A. item. Additionally, the experts provided commentary on the ordering of items to improve readability and simplicity for the finalized survey. The recommendations for item placement within the survey helped to reorder items such that items with similar response options were bunched together for simplicity for respondents.

Final Content Expert Approval

After using the controlled feedback from Round 2, we sent a revised survey to experts for final approval. At this time, with only one item needing relevance revision and three for clarity (two which have needed clarification since Round 1), reviewers no longer needed to rate each item. Rather, we asked the experts to either 'approve' or 'disapprove' of the finalized survey and provide any last minute feedback on the survey. In general, experts made recommendations on spelling, grammar, and overall formatting of the survey (e.g., placement of response options in relation to the item and stem). Every reviewer 'approved' of the finalized survey, including the expert who missed Round 2. We used the comments received to make the necessary grammatical, spelling, and formatting changes. Experts recommended that we change the response scale for five frequency items from the 'strongly disagree to strongly agree' scale to a frequency scale 'never to always'. This made sense since the five items were referring to frequency of assessment. Originally, the items were measuring one's perception of their frequency of assessment. This was a welcome change as the new response scale is better in line with our research questions. One reviewer continued to make changes to the item 4 clarity and

relevance by helping to categorize the listed motor skills into the ‘locomotor’, ‘manipulative’, and non-locomotor’ subcategories.

Inservice Physical Education Teacher Criteria

All inservice elementary physical education teachers (n = 11) were certified physical education teachers currently working in district, with only one retired. Sixty-four percent possessed 11 or more years of teaching experience at the elementary school level. Eighty-two percent of the teachers held a graduate degree and received at least one course on assessment in physical education. Regarding current use of assessment in teaching, more than half declared they frequently use or have used a variety of formal and informal assessments, including diagnostic, formative, and summative assessments. All teachers reported to have at least used 1-2 different types of assessment, including the diagnostic, formative, and summative assessments. Three of the teachers had published a journal article of varying degrees (1- practitioner, 1- low-tier research, 1- top-tier research). Eighty-two percent of teachers had participated in a statewide assessment project in their respective states. The majority of inservice elementary physical education teachers were originally selected for participation based on their good standing with the university as high quality student teaching mentors and/or their utilization of assessment in physical education. Table 2 describes inservice elementary teacher responses on the ‘Expert Rater Table’.

Inservice Teacher Round

Inservice teachers received a version of the survey after all of the content experts approved with a total of 64 items to rate on relevance and clarity, with the option to leave comments about the items. The survey included two items for assessment behavior, six

items for perceived assessment importance, 18 items for frequency, two items for assessment environment, four items about assessment likelihood, 15 perception items, 13 learning opportunity items, and four E.S.S.A. items (See Table 3.3). Additionally, the survey included instructions and examples for the inservice teachers to use while filling out the Delphi along with a set of definitions and key terms they would see throughout the survey to reduce any confusion.

Relevance and Clarity

Only two items were flagged for relevance and clarity. Perception item 71 was flagged for relevance consensus (M = 2.70, SD = 1.25). This item refers to student influence on teacher behavior (i.e., "My students' perceptions of motor skill assessment influence the frequency for which I assess motor skills" on a six-point Likert scale from Strongly Disagree to Strongly Agree). Similar to the relevance consensus, perception item 71 was also flagged for clarity consensus (M = 2.90, SD = .99). Inservice teachers recommended this item was too wordy, confusing, and had never crossed their mind as an influence over their assessment behavior. For these reasons, this item was eliminated.

Frequency item 14 was rated 3.0 (SD = 1.25) for relevance and clarity. This item could not be deleted because it was an item designed to answer part of the research question. Therefore, we defaulted to content expert consensus and decided not to delete this item. Item 14 asked about the use of the Test of Gross Motor Development -2 or -3 (TGMD-2, -3) as a motor skill assessment. Low inservice consensus ratings for this item are a finding in itself, indicating that the tool likely is not one used among this population of teachers.

Reducing and Adding Items

Item 71 was eliminated due to low relevance and clarity consensus. We reviewed all inservice teacher comments and recommendations for item changes regardless of consensus rating for relevance and clarity. This helped us gain greater understanding for survey implementation with the future population of inservice teachers. The recommendations and comments served to support the research team in any change of wording, phrasing, or readability of items in the final version.

Final Research Team Round

Inservice teachers' overall consensus ratings across all items for relevance (M = 3.8, SD = .25) and clarity (M = 3.76, SD = .24) were high, with only one item recommended for removal. Due to the nature of inservice teachers rating very highly with limited comments and recommendations for change on other items, we decided to forgo a second round with teachers. Instead, as a research team, we took a final glance at the survey items to ensure statistical utility. Our goal was to ensure that the data from this survey would inform our research questions, use response scales that can be measured for reliability in future studies, and ensure items were not convoluted or double-barreled. With this said, we decided to split the two assessment behavior items into four items to be more specific within each item (i.e., teach and assess both locomotor and manipulative skills). We reduced assessment importance into four items rather than six, by focusing on the teaching and assessing of 'locomotor' and 'manipulative' skills, removing non-locomotor skills, which is more in line with our research questions. We added ten frequency items by splitting one multiple choice item into several items and turning several resource examples into several items. The splitting of these items would better

inform our statistical analyses for reliability. We added one perception item about class sizes influencing assessment behavior to not only identify if this would influence frequency of assessment but also quality of assessment. We also added one E.S.S.A. item to address awareness of E.S.S.A. before asking whether or not the teacher's district receives and utilizes the funding. By condensing the items, we eliminated two importance items. We also eliminated item five, which was consistently flagged for consensus throughout rounds for confusion. Item 71 was eliminated by the inservice teachers, and we eliminated learning opportunity items 45 and 91, due to teacher comments about confusion and relevance. Item 45 initially asked about professional development opportunities (listing one-day versus a series of professional development opportunities) in a multiple choice format, so we created an individual item for each. Item 91 mentioned teachers using district-wide learning communities as a method for learning opportunities about motor skill assessment. Teacher comments regarding Item 91 was that they do not know what these district-learning communities are. Therefore, this item was deleted.

The research team produced a survey that did not change item content but only format by splitting or removing items to ensure statistical utility for reliability testing. After adding and removing items, our survey had a total of 72 items comprised of four behavior items, four importance items, 27 frequency items, two environment items, four likelihood items, 15 perception items, 11 learning opportunity items, and five E.S.S.A. items (See Table 3.3). To ensure the changes that the research team made did not disrupt the overall content validity of the survey, we sent the survey back out to our content experts for their ultimate final approval.

Final Approval from Content Experts

Seventy-nine percent (n = 11) of content experts responded to the final approval round of the Delphi process, all of whom approved of the final version. Three content experts did not respond to final approval. Six of the 11 experts provided additional comments and recommendations to the survey. Most of the expert comments were working to continue clarifying item 4 (the list of motor skills). Using content expert feedback, we added one final frequency item (15b), to differentiate where teachers receive their motor skill assessments, either from their district or from their state. This is an important item to add since there are some states that utilize state-mandated assessments and others that do not. A final change was made to move perception item 78 to the frequency scale since it utilizes frequency response options for statistical purposes. After the entire modified Delphi process was over, we thanked each of the content experts and inservice teachers for their participation through email and by sending a 'thank you' card in the mail.

Final Survey

The final survey consists of 73 items, the same number of items as the 'Final Approval from Content Experts' version except that the frequency item count changed from 27 to 28. Table 3.3 represents the overall changes to the survey by phase and divided by item theme.

Survey Subscales

Due to the large volume of survey items housed within this survey, we asked the content experts to consider the survey being broken into several subscales, by theme (e.g., frequency, perception, learning opportunities). Being that all of the items earned

content expert consensus, thus obtaining content validity, we only needed approval that the content experts could consider separately disseminating subscales of this survey due to its overall length (i.e., 73 items). It is understood in the survey literature that the researcher should only include the number of items in a survey that they believe the respondent will spend answering, thus improving survey response rate (Johnson & Morgan, 2016). All of the experts who responded to the Final Approval round (79%) approved of using this survey in subscales.

Discussion

The purpose of this modified Delphi process was to explore content validity for survey items surrounding the topic of assessment of motor skills in elementary physical education to be used in the future with elementary physical education teachers.

Strengths

First, the inclusion of both content experts in the field of assessment in physical education and currently practicing inservice elementary physical education teachers serve as strengths to this study. The goal was to ensure that the items within the survey represented appropriate conceptions and types of assessment, were relevant to assessment behavior that takes place in the elementary physical education setting, and specifically addressed the formal assessment of motor skills. Additionally, the multiple revisions to the survey based on the Delphi consensus and content expert and inservice teacher comments served as a strength and helped to holistically develop the content presented within the survey items.

Assessment in physical education remains a practice that is unmonitored (i.e., accountability; van der Mars et al., 2018) and is often forgone (Lander et al., 2016, 2017;

Leirhaug & MacPhail, 2015; Lund & Veal, 2008). The measurement of preservice and inservice teacher perceptions and frequency of assessment practices, especially in regard to motor skills, is limited, if not absent, in the published literature. Within the extant literature, there is very limited inquiry that features a variety of methodologies to measure frequency and perception of assessment. Thus, it was difficult to compare previous findings across studies. The development of a survey that measures frequency and perception with sound psychometric properties might help with generalizability and an increase in studying different populations of teachers.

Limitations

First, not all experts responded to Round 2 of the Delphi process; however, they all responded to the subsequent final approval round and the reviewer who skipped Round 2 provided extra commentary at the first final approval round of the Expert Phase (before going to inservice teachers). Second, although consensus of survey items was met at appropriate levels, representing content validity, we did not assess instrument reliability. Future research should consider the reliability of survey items within and across themes (or subscales).

Implications for Practice

The survey items within the Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey demonstrate content validity and can be used to measure inservice physical education teacher assessment behaviors related to motor skills. In order to interpret results, the psychometric properties should be determined. This survey can be utilized in totality to measure overall assessment behavior or by subscale to measure certain elements of motor skill assessment behavior. The data gleaned from using this

survey tool can be utilized across multiple populations of teachers and therefore improve generalizability of findings. Interpretations of the data drawn from this survey tool can help to inform teacher training, both in physical education teacher education programs and through professional development for inservice teachers.

Conclusion

The modified Delphi method proved to be a successful tool to determine content validity on survey items by using a purposive sample of content experts and inservice elementary physical education teachers. The survey went through various iterations, improving with each round. By the end of the modified Delphi process, both content experts and inservice teachers met consensus ratings for item relevance and clarity, thus ensuring content validity of survey items. Therefore, the survey items and subscales can now be used to measure inservice elementary physical education teacher frequency and perception of motor skill assessment. Future research should investigate the reliability and psychometrics of this survey tool.

Table 3.1. Content Expert (n = 14) Criteria

Criteria	Criteria Options	Expert (*n* = 14) Frequencies
Content Knowledge	Taken 1 University Course	11
	Taught 1 University Course	6
	Taught Course and/or published textbook	6
	Currently teaching course and/or published(es) textbook(s)	5
	Taught professional development to K-12	1
Delivered Assessment in teaching	Used 1-2 of the following types of assessment: formal, informal, Diagnostic/ Formative/ Summative	7
	Used a variety of formal, informal, Diagnostic/ Formative/ Summative	7
	Frequently uses a variety of formal, informal, Diagnostic/ Formative/ Summative	5
	Consistently uses a variety of formal, informal, Diagnostic/ Formative/ Summative	11
Published	Have not published	1
	In JOPERD or Strategies	7
	Nationally or internationally in lower-tier journal	5
	Nationally or internationally in higher-tier journal	9
Assisted in statewide assessment development	Yes	9
	No	5
Highest Completed Education	Bachelors	0
	Graduate	0
	Doctoral	14
K-12 Teaching Experience	No K-12	1
	Student Teaching only	0
	1-5 years	4
	6-10 years	6
	11+ years	3
Current Employment	District	1
	District & University	0
	University/College	10
	Education Contractor	0
	Retired	3

Current Role at Employment	Certified physical education teacher	0
	Graduate assistant in masters' degree	0
	Graduate assistant in doctoral degree	0
	Professorship	
	Assistant Professor	1
	Associate Professor	2
	Full Professor	7
	Professor Emeritus	3
	Retired Teacher	1
	District Coordinator or Director in Health and Physical Education	1

Table 3.2. Inservice Teacher (n = 11) Expert Criteria

Criteria	Criteria Options	Inservice Teacher (*n* = 11) Frequencies
Content Knowledge	Taken 1 University Course	11
	Taught 1 University Course	0
	Taught Course and/or published textbook	0
	Currently teaching course and/or published(es) textbook(s)	0
	Taught professional development to K-12	0
Delivered Assessment in teaching	Used 1-2 of the following types of assessment: formal, informal, Diagnostic/ Formative/ Summative	6
	Used a variety of formal, informal, Diagnostic/ Formative/ Summative	4
	Frequently uses a variety of formal, informal, Diagnostic/ Formative/ Summative	6
	Consistently uses a variety of formal, informal, Diagnostic/ Formative/ Summative	5
Published	Have not published	9
	In JOPERD or Strategies	1
	Nationally or internationally in lower-tier journal	1
	Nationally or internationally in higher-tier journal	1
Assisted in statewide assessment development	Yes	2
	No	9
Highest Completed Education	Bachelors	2
	Graduate	9
	Doctoral	0
K-12 Teaching Experience	No K-12	0
	Student Teaching only	0
	1-5 years	1
	6-10 years	3
	11+ years	7
Current Employment	District	10
	District & University	0
	University/College	0

	Education Contractor	0
	Retired	1
Current Role at Employment	Certified physical education teacher	11
	Graduate assistant in masters' degree	0
	Graduate assistant in doctoral degree	0
	Professorship	
	Assistant Professor	0
	Associate Professor	0
	Full Professor	0
	Professor Emeritus	0
	Retired Teacher	0
	District Coordinator or Director in Health and Physical Education	0
Years' Experience Teaching at the Elementary Level	1-5 years	1
	6-10 years	3
	11+ years	7

Table 3.3. Complete Overview of Frequency Items by Phase and Theme

C. Exp. Round 1	**Item N**	**C. Exp. Round 2**	**Item N**	**Tch. Round**	**Item N**	**Final Rsch. Team and C. Exp. Approval**	**Item N**	**Final Survey**	**Item N**
Total	50	**Total**	72	**Total**	64	**Total**	72	**Total**	73
Behav.	1	**Behav.**	2	**Behav.**	2	**Behav.**	4	**Behav.**	4
Imp.	2	**Imp.**	4	**Imp.**	6	**Imp.**	4	**Imp.**	4
Freq.	20	**Freq.**	27	**Freq.**	18	**Freq.**	27	**Freq.**	28
				Env.	2	**Env.**	2	**Env.**	2
Likely	3	**Likely**	4	**Likely**	4	**Likely**	4	**Likely**	4
Perc.	17	**Perc.**	23	**Perc.**	15	**Perc.**	15	**Perc.**	14
LO	4	**LO**	9	**LO**	13	**LO**	11	**LO**	11
ESSA	3	**ESSA**	3	**ESSA**	4	**ESSA**	5	**ESSA**	5

Note: C. Exp. = content expert; Tch. = teacher; Rsch. = research; Behav. = assessment behavior items, Imp. = assessment importance items, Freq. = frequency items, Env. = assessment environment items, Likely = likelihood items, Perc. = perception items, LO = learning opportunity items, and ESSA = E.S.S.A. items.

To what extent do you agree with the following statement? Response option for each item (1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree)				
Item #	**Item Content**	**Relevance (check box)**	**Clarity (check box)**	**Comments**
Item 2	Locomotor skills (i.e., run, jump, leap) should be taught in elementary school physical education.	☐ Very Irrelevant ☐ Irrelevant ☐ Relevant ☐ Very Relevant	☐ Very Unclear ☐ Unclear ☐ Clear ☐ Very Clear	

Figure 3.1. Sample Expert Rater Table for One 'Importance' Item

CHAPTER 4

STUDY 2: PSYCHOMETRIC PROPERTIES OF THE ELEMENTARY PHYSICAL EDUCATION TEACHER PERCEPTIONS OF MOTOR SKILL ASSESSMENT SURVEY SUBSCALE

Introduction

Assessment in physical education is understood to be a pedagogical skill often omitted due to a wide variety of perceived barriers (Morgan & Hansen, 2008; Penney, 2012; Stiggins, 1997). Physical education teachers perceive major barriers to assessment, such as a lack of time, space, equipment, knowledge, and value (DinanThompson & Penney, 2015; Georgakis, Wilson, & Evans, 2015; Kneer, 1986; Lander et al., 2015, 2016, 2017; Leirhaug & MacPhail, 2015; MacPhail & Halbert, 2010; Matanin & Tannehill, 1994; Michael, Webster, Patterson, Laguna, & Sherman, 2016; Veal, 1988). Other physical education teachers believe there are too many students to be able to assess in a short amount of time, with limited educational training to assess, and limited accountability in district (Hensley, 1990; Kneer, 1986; Veal 1988). To overcome these barriers, physical education teachers prefer assessments that can be used in authentic settings (Haynes & Miller, 2015; James, Griffin, & Dodds, 2009; Lander et al., 2016; MacPhail & Halbert, 2010; Mintah, 2003; Ni Chroinin & Cosgrave, 2013; Patton & Griffin, 2008; Richard, Godbout, Tousignant, & Grehaigne, 1999) but also recognize a strong need for additional assessment training to be successful (Annerstedt & Larsson, 2010; Lander, 2015, 2017; Lund & Veal, 2008; Michael et al., 2016; Slingerland et al.,

2017). Although each perception study adds its own element to the physical education literature, the limited ability to generalize across studies and samples is difficult.

The methodologies used to measure physical education teacher perceptions of assessment have varied (i.e., qualitative, quantitative, and mixed-method), with no consistent measurement tool used across any perception studies. The variables measured within each study differ based on the study's specific needs, thus limiting the ability to compare one sample of physical education teachers to the next. Additionally, many of the physical education perception of assessment studies are conducted outside of the United States of America (Fisher et al., in preparation), using different physical education standards and content, and with teachers in varying levels of education (e.g., middle school or high school). To our knowledge, not one of the perception studies measured elementary teacher perceptions of assessment or the assessment of motor skills. Currently, there is no psychometrically sound measure to compare teacher perceptions of assessment in any physical education content or topic area.

The Society of Health and Physical Educators (SHAPE America) Research Council recently published a national call for teachers to measure student motor skills in the United States of America (Castelli & van der Mars, 2018). Student's basic motor skills are primarily taught at the elementary level (SHAPE America, 2013) providing students with the fundamental skills needed to build into more complex movement patterns as they grow older (Logan et al., 2015; Lubans et al., 2010). Therefore, elementary physical education teachers should be targeted to measure formal assessment behavior (i.e., collection of evidence; Lund & Tannehill, 2005) of student motor skills. With a lack of assessment perception research studies in the United States of America

taking place at the elementary school level, and with thoughts that assessment perception likely motivates assessment behavior (Lander et al., 2015, 2016; Leirhaug & MacPhail, 2015), we feel it is important to measure elementary physical education teacher perceptions of motor skill assessment.

The purpose of this study was to measure the reliability of the Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey subscale using a national sample of elementary physical education teachers. Additionally, investigated the factor structure of items within the perception scale to determine if there are underlying factors that make up teacher perceptions of motor skill assessment. More specifically, our research question was "What are the psychometric properties of the Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey subscale using a population of inservice elementary physical education teachers in the United States of America?". We hypothesized that there will be several factors within the perception subscale that align with the themes from the literature (i.e., lack of support, lack of time, class sizes too large, etc.; Fisher et al., in preparation). Additionally, we hypothesized that our survey items will produce reliable results using our sample.

Methods

Design

This study is a descriptive analytic and cross-sectional study. We used the perception survey subscale that was developed, and which demonstrated content validity, in Study 1.

Sample

Participants included a voluntary national sample of certified public school elementary physical education teachers (N = 445) who responded to the survey. Individual (personal) demographic item responses can be found in Table 4.1, while general demographic and school context item responses are reported in Table 4.2. A total of 296 participants (Female, n = 233, 78.7%) responded to personal demographic items with the majority of elementary physical education teachers between the ages of 31-35 (n = 54, 18.2%) and ages 36-40 (n = 56, 18.9%) and white (n = 280, 94.6%), not Hispanic or Latino (n = 274, 92.6%). Of the 296 that responded to individual demographic items, the majority of elementary physical education teachers had between 1-5 (n = 58, 19.6%), 6-10 (n = 58, 19.6%), 11-15 (n = 56, 18.9%) years of teaching experience in physical education.

Sample Recruitment

The lead researcher started with a premade listserv of 441 elementary physical education teachers across the United States of America that was developed for a different study (Webster et al., 2019). To create the listserv, Webster and colleagues (2019) obtained a list of public school districts from the National Center for Educational Statistics and began manually searching physical education teacher school district email addresses from their school webpages. With a focus on elementary physical education teachers, the lead researcher employed this same strategy to obtain more elementary physical education teacher email addresses per state across the nation. Using the 441 email addresses from the initial premade listserv, the lead researcher manually collected an additional 2,078 more from district webpages. The state of Hawaii protects their

teacher's school's email addresses from the public domain; therefore, Hawaii was the only state for which we did not have representation in our stratified random sample. We used a total of 2,519 teacher's district email addresses for survey dissemination.

To gain a wider reach of elementary physical education teachers, the lead researcher investigated other methods of recruitment, such as popular Facebook groups (e.g. Standards-based Physical Education, SHAPE America, PE Central, Elementary PE Teachers, PE Teacher Community, etc.) and Twitter accounts (each state SHAPE organization with a Twitter account, PE Central, PhysEd, etc.) for elementary physical education teachers. Additionally, the lead researcher reached out to 322 SHAPE America executives (found on state and national webpages) and 43 state physical education representatives (others were not available on the state websites) asking for these individuals to forward the survey to their own listserv of elementary physical education teachers. Most often, due to confidentiality, these individuals were not able to provide the lead researcher with a list of email addresses to add to the research listserv but were able to forward the survey invitation email on to their teachers. Not all of the state representatives were diligent in providing the exact number of elementary physical education teachers that were included in their forwarded emails. Although this addition was beneficial in improving the study sample size, it serves as a limitation because it made it more difficult to collect a survey response rate. Teachers were incentivized to participate in this survey by selecting the option to be entered into a raffle for a $50 Amazon gift card.

Instrumentation

The survey instrument implemented in this study was intended to measure elementary physical education teacher perceptions of motor skill assessment. The survey items included in this study represent the 'perception subscale' of a larger 'assessment behavior' survey (see Study 1). All of the items within the 'assessment behavior' survey and this 'perception' subscale have demonstrated content validity (see Study 1). The survey is formatted electronically, collected, and managed by REDCap software (Harris et al., 2009; Harris et al., 2019) hosted by the University of South Carolina.

The perception survey subscale ($N = 21$) includes four behavior items, four importance items, and 13 perception items. The importance items were included within the perception scale measure perceived importance of teaching and assessing motor skills. Additionally, both scales use the same response options (e.g., six-point Likert scale from strongly disagree to strongly agree). See the full survey in Appendix A.

Assessment behavior is measured using a dichotomous response scale (yes, no) with a focus on formal assessment behavior. An item example includes "Do you formally assess (written/documented) your students' ability to perform locomotor skills (i.e., skip, jump, run) in your physical education classes?" and "Do you formally assess (written/documented) your students' ability to perform manipulative skills (i.e., throw, catch, strike) in your physical education classes?".

Example items measuring the perceived importance of assessment of motor skills includes "Locomotor skills (e.g., run, jump, leap) should be taught in elementary school physical education" and "Locomotor skills (e.g., run, jump, leap) should be formally assessed (written/documented) in elementary physical education". Importance items fall

on a six-point Likert-type scale (1 = strongly disagree, 3 = somewhat disagree, 4 = somewhat agree, 6 = strongly agree).

The item stem for perception related questions is: “To what extent do you agree that the following statements influence your use of formal motor skill assessment?” with subsequent item examples including “The time needed to formally assess motor skills negatively impacts the opportunity for my students to learn motor skills in class” or “I typically have enough time to formally assess motor skills”. Perception items utilize a six-point Likert-type scale (1 = strongly disagree, 3 = somewhat disagree, 4 = somewhat agree, 6 = strongly agree).

The survey also includes demographic and school context questions, including teacher sex, years of teaching experience, location of school (i.e., rural vs. suburban), Title I school, etc. (see Tables 4.1 and 4.2). Additionally, there is an option to provide an email address for future correspondence to participate in a qualitative interview at a later date for individuals who wish to explain in greater detail their assessment perceptions. Due to our understanding and importance of teacher’s time, the survey is constructed such that it can be completed in several increments, if needed, with the ability to save and finish the survey later. The survey should take no longer than 10 minutes to complete.

Procedures

This study is the second phase of a larger study (Study 1). During phase one, we developed survey items surrounding the topic of motor skill assessment behavior and obtained content validity using a sample of content experts and inservice elementary physical education teachers. This study consisted of obtaining the psychometric properties (i.e., reliability and factor structure) of only the perception subscale using a

national sample of elementary physical education teachers. Prior to data collection, we acquired Institutional Review Board approval through the University of South Carolina. The lead researcher began the collection of elementary physical education teacher email addresses. This process spanned several months, including time when the survey was already publicly accessible. Prior to dissemination, we converted the survey items from the modified Delphi study (Study 1) to the online survey software, REDCap (Harris et al., 2009; Harris et al., 2019), where it was disseminated. We created an email invitation to participate that briefly explained the study and described the optional incentive (raffle for a $50 Amazon gift card). Every Monday for three months, we sent email reminders for individuals to participate and/or thanking them for their participation. Additionally, we posted a shortened invitation and weblink to the survey to the identified Facebook groups and Twitter accounts. We sent formal emails with the invitation and survey weblinks to the SHAPE executives and state physical education representatives asking for those individuals to either provide us with the list of email addresses or to forward the invitation and weblink onto their listserv.

Data Analysis

Prior to data analysis, we checked the data for direction of items, normality, linearity, factorability, miscodes, the number of cases, and that all assumptions are met to run our statistical analysis (Bandalos, 2018). All statistical analyses were conducted using the Statistical Package for the Social Sciences (Version 26; SPSS Inc., Armonk, NY: IBM Corp.).

Data Screening and Preparation

Descriptive Statistics

We ran descriptive statistics (mean and standard deviations) on all motor skill assessment behavior items, perception items, and demographic items. We also checked skewness and kurtosis on perception items to test for normality, linearity, and to make sure the data was clean. Specifically, with survey design, it is important to check for skewness and kurtosis to determine if responses to certain items will distort the data. Bandalos (2018) describes skewness (positive or negative) as "the degree to which an item's distribution deviates from symmetry" within a distribution curve with acceptable ranges within |2| (p. 132). Kurtosis concerns the 'peakedness' or 'tailedness' of an item distribution in relation to the way individuals respond to items (e.g., most participants choosing the same response option would result in high positive kurtosis; Bandalos, 2018). Recommended values for kurtosis typically fall within |2| (Bandalos, 2018), with some researchers considering a more tolerant kurtosis value of |7| (Kline, 2005).

Homogeneity

Survey items were tested for homogeneity by running an item-total correlation. A corrected item-total correlation measures one item's correlation with the sum of all of the other items, besides itself, to ensure the correlation is not inflated (Bandelos, 2018). Higher correlations indicate the item's ability to discriminate between different groups of people (e.g., teachers who assess versus teachers who do not assess), where lower correlations indicate an item is not good at differentiating between groups (Bandalos, 2018).

Internal Consistency Reliability

We assessed internal consistency of the items using Cronbach's coefficient alpha, which is a commonly used reliability measure for survey scales (Bandalos, 2018). Coefficient alpha is based on interitem correlations, therefore, the correlations of each item to every other item on the scale. Coefficient alpha is a correlation statistic that demonstrates consistency with other items on a scale with higher coefficient alpha demonstrating better internal consistency (Bandalos, 2018). An important consideration in determining coefficient alpha is to evaluate the 'alpha-if-deleted' statistic which represents the overall scale reliability if each item were removed. If the reliability is predicted to improve without an item (if deleted), the item is likely detrimental to internal consistency, and researchers should consider removing the item (Bandalos, 2018).

Exploratory Factor Analysis

Factor analysis is a statistical procedure used to determine how items relate to an overarching construct (i.e., perception of motor skill assessment) (Bandalos, 2018). We conducted an Exploratory Factor Analysis (EFA) to determine the relationships between survey items and the associated constructs (motor skill assessment perceptions). Through the use of EFA, we were able to determine the psychometric properties (i.e., preliminary internal consistency and factor structure) of the perception survey tool within our sample by evaluating factor loadings (pattern and structure matrix), item communalities, and factor correlations (Bandalos, 2018). Only Likert-type items from the perception scale are included in factor analysis as factor analysis requires continuous or interval data (Bandalos, 2018). Exploratory factor analysis utilizes item correlations and covariances to create factors, or groups, of items that are similar with respect to the patterns for which

participants respond to items. Specifically, items that have higher inter-item correlations will likely load together in the same factor (Bandelos, 2018). Pearson product-moment correlations are used for EFA estimation. Therefore, it is important that the data does not violate the assumptions of continuous and linear data (Bandalos, 2018). Under classical test theory, survey scales with five or more categories can be considered continuous data (Bandalos, 2018).

Exploratory factor analysis often requires between five, 10, and 20 times the number of respondents per survey item, or for factor structures with mediocre communality values, a sample size of approximately 300 respondents are needed (Bandalos, 2018). The perception survey consisted of 17 six-point Likert type items. Therefore, based on Bandalos (2018) recommendation, our sample size of 445 was sufficient to run an EFA to determine factor structure. Additionally, to test for sampling adequacy, we used the Kaiser-Meyer-Olkin (KMO). Specifically, the KMO looks to ensure the variability among item responses is enough for the items to be factored. Kaiser (1974) recommends anything above .70 ("mediocre", moderate or average) as acceptable.

We used principal axis factoring to reduce the data with an oblique-oblimin rotation using 'motor skill assessment perceptions' as the latent variable. We expected that the factors would relate to each other within the overall construct (motor skill assessment perceptions). An oblique rotation method would demonstrate (with correlations = 0) if there was no correlation and would default to the orthogonal rotation, if necessary. Finally, oblimin rotation was used because it is the most common rotation method used for oblique rotations (Bandalos, 2018). We reverse coded three items so that all items result in a positive direction before analysis (Bandalos, 2018). We determined

the number of factors (using a variety of factor structures), number of item loadings per factor, variance explained by each factor, and the range of variance explained by each item within a factor. To determine the number of factors to retain, we looked for eigenvalues greater than 1 (K1 criterion) from the factor analysis output and a Scree plot for factors above the "elbow" (Bandalos, 2018). To determine a final solution, we first looked at factor loading values (lambda) above .4 for an item within a factor and each item's communality (item shared variance with all factors) above .2. Next, we looked for simple structure, such that each item primarily loads (lambda factor loading value) to one factor with under .3 loading value for any other factor (crossloading). Subsequently, we looked at the percent of variance explained by each factor to ensure there was good spread between each factor. Additionally, it checked if the factor structure was over-factored (when there is cross-loading and not primary loadings or only one to two items per factor). Finally, we checked to ensure the factor structure solution matched the theory underpinning the measured construct. This involved determining whether the factor structure was interpretable based on where items were loading and if that made sense theoretically (DiStefano & Dombrowski, 2006; Henson & Roberts, 2006).

The results from the EFA provide evidence as to whether the survey items need to be revised and in what ways (i.e., rewording, removal from survey, or changing from positive to negative direction, etc.) to better represent the overall construct being measured. In summary, within this motor skill assessment perceptions survey, we were looking to see how the items we asked help to explain the overall construct of perception of motor skill assessment using the population of elementary physical education teachers in our sample.

Results

Descriptive Statistics

Sample Demographics

A total of 445 elementary physical education teachers responded to this survey. Tables 4.1 and 4.2 describe the individual (personal) demographics and school context of the sample. The majority of teachers were from South Carolina (*n* = 41; 9.2%), Montana (*n* = 34; 7.6%), Louisiana (*n* = 31; 7%), Illinois (*n* = 25; 5.6%), Maryland (*n* = 25; 5.6%), and Georgia (*n* = 23; 5.2%). Most teachers taught 100% general physical education classes (*n* = 319; 71.2%) with 100 teachers (22.5%) teaching about 75% general physical education and 25% of their classes being adapted physical education classes. Typical physical education departments were made up of only one elementary physical education teacher (*n* = 193; 43.4%) and most others included two faculty (*n* = 144; 32.4%). Teachers reported that a typical physical education class duration lasted from 26-30 minutes (*n* = 99; 22.2%) to 41-45 minutes (*n* = 151; 33.9%). Most physical education classes were comprised of 21-25 students (*n* = 178; 40%) and 26-30 students per class (*n* = 105; 23.6%). Twelve percent (*n* = 56) of teachers reported having 16-20 students in each class, while 11.2% (*n* = 50) reported having more than 46 students in each physical education class. Physical education teachers typically saw their students between one (*n* = 130; 29.2%) and two times (*n* = 190; 42.7%) per week.

Most teachers taught in Title I (*n* = 236, 53%) public schools (*n* = 396, 89%) with more than 50% of their students on free or reduced lunch (*n* = 197, 44%). Forty percent of teachers (*n* = 176) do not teach in Title I schools. School location was pretty evenly split between urban (*n* = 166, 26%), suburban (*n* = 181, 41%), and rural (*n* = 141, 32%).

Most schools were made up of 251-500 students (*n* = 181, 41%) or 501-750 students (*n* = 134, 30%).

We asked teachers if there was a physical education district supervisor in place at their school who is responsible for the overall goals of assessment in physical education. Sixty four percent (*n* = 288) of teachers declared that there is no district supervisor and 29% (*n* = 128) said there is a supervisor in place. Twentynine teachers (6.5%) did not know if there was anyone in their district. When asked if their elementary school has a policy in place for assessment in physical education, 66.3% (*n* = 295) said 'no', 27.4% (*n* = 122) said 'yes', and the rest did not know. Teachers responded similarly when asked about district policy (no = 230, 51.7%; yes = 167, 37.5%; I do not know = 48, 10.8%). However, when we asked teachers about a state policy for assessment, teachers responded in a more distributed way. The majority of teachers (*n* = 162, 36.4%) did not know if their state had a policy for assessment, while 35.3% (*n* = 157) said 'yes' and 28.3% (*n* = 126) responded 'no' they do not have a statewide assessment policy. Favorably, teachers believe that their building principal is very supportive (*n* = 169, 38%) or supportive (*n* = 147, 33%) in regard to physical education. Twenty one percent (*n* = 93) of teachers who responded to this survey believed that their building principal's support for physical education is 'mediocre'.

Assessment Behavior

Almost all elementary physical education teachers declared that they do teach locomotor (*n* = 440, 98.9%) and manipulative (*n* = 444, 99.8%) skills. However, only sixty percent (*n* = 269) and 58.7% (*n* = 261) report assessing locomotor and manipulative skills in elementary physical education, respectively.

Perception Items

See Table 4.3 for descriptive statistics for perception items. First, we checked for distribution normality (skewness and kurtosis) and all data met the assumptions required for statistical analysis except for two importance items. Importance item 2 (teach locomotor skills) and importance item 53 (teach manipulative skills) were both negatively skewed and highly kurtotic. Item 2 (*M* = 5.59, *SD* = 1.20) presented with -3.30 skewness (*se* = .12) and 9.76 kurtosis (*se* = .23). Item 53 (*M* = 5.57, *SD* = 1.20) demonstrated -3.23 skewness (*se* = .12) and 9.41 kurtosis (*se* = .23). Skewness and kurtosis for these items makes sense being that the focus of elementary physical education is to teach locomotor and manipulative skills to children, and therefore teachers responded favorably to the importance of teaching these skills. Therefore, we included these items within our analysis, though the reader should be aware of this deviation from normality.

Homogeneity

Corrected item-total correlations are used to discriminate between different groups of people who respond to the survey (Bandalos, 2018). Item-total correlations explain the correlation between an item and the total scale (Bandalos, 2018). Higher correlations indicate the item's ability to discriminate between teachers who assess versus teachers who do not. Lower correlations indicate an item's inability to make this differentiation, thus providing less utility as a survey item representing the measured construct (Bandalos, 2018). Low correlations should be flagged because this demonstrates that an item is not related to other items in the survey scale (Bandalos, 2018). Two importance items (items 2 and 53) demonstrated very low corrected item-

total correlations (r = .13 and r = .14, respectively). Therefore, these items are flagged as items with low discrimination power. Other item's item-total correlations ranged from r = .21 to .59 demonstrating low to moderate correlations and an ability to better discriminate between groups of respondents. The low item-total correlation demonstrates a lack of discrimination power within the survey. Items 2 and 53 two items were also skewed and highly kurtotic, indicating that these items are likely detrimental to keep in the survey.

Internal Consistency Reliability

Interitem Correlations

Internal consistency reliability is partially based on interitem correlations. Any correlations that are 'too high' should be flagged as items that might be representing the same construct or idea (Bandalos, 2018). All of the correlations that were flagged for a high correlation in this survey is likely due to item wording being very similar between items. Importance items 2 and 53 were flagged for high correlation (r = .86), likely because the item asks the extent to which teachers believe teaching locomotor (item 2) and manipulative (item 53) skills are important in elementary physical education. Almost identically, items 81 and 83 measure the extent to which teachers believe assessing locomotor (item 81) and manipulative (item 83) skills are important in elementary physical education. Item 27 and 27b has high correlations (r = .91) because both items are asking one's perception on the amount of class time influencing quality (item 27) and frequency (item 27b) of assessment. The same is true for items 72 and 43 (r = .92), which also only have one word difference between the items. Items 72 and 43 ask the extent to which the teacher believes the field should standardize assessment across the state (item

72) and country (item 43). Due to high interitem correlation values, these items will likely be factored together.

Internal Consistency Reliability

Reliability of the survey scale is determined by examining item-total correlations and Cronbach's coefficient alpha.

Cronbach's Coefficient Alpha

We used Cronbach's coefficient alpha as a measure of internal consistency reliability for the scale. Furthermore, we evaluated the 'alpha-if-deleted' to determine which items were detrimental to overall reliability of the survey scale. We found that two importance items (item 2 and 53) were impacting overall reliability, likely due to their high interitem correlation. Reliability would improve from .80 to .81 if items 2 and 53 were flagged for removal before running an exploratory factor analysis.

Removal of Items

Items 2 and 53 are two items that were not performing well with other items within the perception scale. Item 2 asks "To what extent do you believe that locomotor skills (e.g., run, jump, leap) should be taught in elementary school physical education?" and item 53 asks "To what extent do you believe that manipulative skills (e.g., throw, catch, kick) should be taught in elementary school physical education?". Both items have a six-point Likert response scale from strongly disagree to strongly agree. Items 2 and 53 were skewed and highly kurtotic, had poor discrimination (corrected item-total correlations), and were highly correlated with each other (interitem correlations), and thus had implications to be removed from further analysis using this survey.

Exploratory Factor Analysis

Prior to running an EFA, it is important to check internal consistency reliability, sampling adequacy, eigenvalues, and the scree plot. After determining how many factors to extract, we were able to run the EFA to determine factor structure with the hopes of finding the most parsimonious and theoretically sound simple structure.

Internal Consistency Reliability

After removing items 2 and 53, item-total correlations improved with all inter-item correlations falling between .2 and .63 demonstrating low to moderate correlations and discrimination power. Additionally, Cronbach's coefficient alpha improved to .82 after removing items 2 and 53.

Kaiser-Meyer-Olkin (KMO) Test

The KMO measures sampling adequacy relating to the variance in responses on the survey. Using the Kaiser (1974) recommendation of average sampling adequacy at .7, our survey proved to have adequate sampling at .73.

Eigenvalues (K1 criterion)

Any factor eigenvalue above one is deemed acceptable for factor retention (Bandalos, 2018). We extracted a five-factor solution each with an eigenvalue over one (Table 4.4).

Scree Plot

The scree plot (Figure 4.1) simply graphs the eigenvalues by factor that have been extracted using this model. The factor representing the bend in the 'elbow' is typically where the researcher looks to determine the number of factors to retain, as long as the

elbow is above the eigenvalue of one. The scree plot for this factor structure displays an extraction of five factors, similar to the eigenvalue (K1) criteria.

Factor Solution

The EFA recommended a five factor solution based on eigenvalues (K1) and the scree plot. Therefore, we evaluated factor loadings within a factor (appropriate >.4) and communalities (appropriate >.2) for each item. We used the structure coefficient matrix to determine the most appropriate factor solution since our factors are correlated (Bandalos, 2018). Also, the structure coefficient matrix made the most sense theoretically when compared to the pattern matrix which is made up of regression weights (used for uncorrelated factors). Using the structure matrix, six items loaded to factor one, two items to factor two, two items to factor three, two items to factor four, and three items to factor five (see Table 4.5). The communalities for all items were deemed appropriate (>.2). Each factor was low to moderately correlated with another, indicating that there is a relationship between each factor measuring teacher perceptions of motor skill assessment (Table 4.6). Theoretically the factors could be labeled as 'feelings of support', 'class size', 'competence', 'perceived importance', and 'standardization'. Bandalos (2018) recommends investigating several factor structures during exploratory factor analysis to ensure that proposed factor structures make sense theoretically with the data. Therefore, we also investigated a three and four factor structure.

The four factor structure (Table 4.7) included all items with no item being excluded due to low loading values. All items loaded primarily to one factor, but several had crossloading values between .3 and .4 with other factors. Each item within this factor structure had an appropriate communality value (>.2). Each factor had low to moderate

correlations with each other factor (Table 4.8). Theoretically, the factors in the four factor structure could be labeled ‘standardization and sharing results’, ‘class size and time’, ‘feelings of support’, and ‘perceived importance’.

The three factor structure (Table 4.9) was the most parsimonious model with all items primarily loading to one factor with several crossloading values. There were no items that were eliminated due to low loading values in this model. Each item had an appropriate communality value (<.2). Each factor had a low to moderate correlation with another factor (Table 4.10) Theoretically, the factors in this structure could be labeled ‘support for motor skill assessment, ‘class size and time’, and ‘feasibility of assessment’.

Simple Structure

Simple structure is important in retaining the number of factors while using EFA (DiStefano & Dombrowski, 2006; Henson & Roberts, 2006). Typically, simple structure is represented by at least three items loading per each factor with minimal crossloading (<.3 factor loading value for other factors; Bandalos, 2018). In each of the factor structures (five, four, and three), there were several instances of crossloading values. Although Bandalos (2018) recommends simple structure, she also does not recommend eliminating items for crossloading until the survey has been replicated with another sample because factor analysis can be unstable with poorly correlated items. Within each of the factor structures, the crossloading values did not exceed .51 (occurrence in the three factor structure) but were flagged if the item had a factor loading value above .3 with another factor. The crossloading values are not ideal but may be considerations for use in a confirmatory factor analysis. Ultimately, we decided to utilize a three factor

structure because this model was the most parsimonious and made the most sense both theoretically and statistically.

Discussion

The purpose of this study was to measure internal consistency reliability and factor structure on the Elementary Physical Education Teacher Perception of Motor Skill Assessment Survey subscale using 445 elementary physical education teachers across the United States of America. Fifteen of the initial seventeen items were used to obtain survey reliability in an acceptable range using our sample. Two of the survey items were removed from the survey during the screening process because the data was not performing well with others on the scale (low variance in the data producing skewness and kurtosis, correlation levels above acceptable). The remaining 15 items were used in an EFA to determine if the survey items were truly measuring the latent construct of perceptions and to what extent the items loaded well with each other creating factors within the perception construct. Our three factor model was selected since the data had a moderate fit to the model and this model made the most theoretical sense. Using this three factor model, the factors could be labeled ‘support for motor skill assessment’, ‘class size and time’, and ‘feasibility of assessment’. Each factor has more than four items per factor with appropriate factor loading and communality values.

Describing the Three Factor Structure

The factors within the three factor structure align with the themes found within the physical education teacher perceptions of assessment literature (see Chapter 2, Literature Review and Fisher et al., in preparation). Each survey item followed the survey stem ‘to what extent do you agree that the following statements influence your use of

formal motor skill assessment' with responses on a six-point Likert scale from strongly disagree to strongly agree. The three factors were created using the EFA and are based on a pattern of participant response to each item. Therefore, it is not discernable whether or not our sample mostly agreed or disagreed to any of the items, but rather, that they responded similarly.

'Support for Motor Skill Assessment'

The first factor, 'support for motor skill assessment', consists of items referring to commonly made recommendations for assessment in physical education. Items include the standardization of assessment across the state and country, the importance of assessing motor skills (both locomotor and manipulative), the perceived importance of sharing assessment results with students, and the belief that assessment is a good use of teacher time in physical education. Therefore, for these items to form a factor, individual participant patterns of response could either agree or disagree with these items themselves as influences over motor skill assessment behavior. Commonly recommended by teachers in physical education is the development of a simple, authentic, valid, and reliable assessment tool that can easily be used in physical education settings (DinanThompson & Penney, 2015; Georgakis et al., 2015, Lander et al., 2017; Leirhaug & MacPhail, 2015). Perhaps the item describing a desire for a standardized assessment tool across the state and country would represent this perceived need and support for assessments within this factor. Although teachers often understand the importance of teaching and assessing motor skills in physical education (Lander et al., 2015; 2016), they often lack the knowledge and skills to do so (Borghouts et al., 2017; Goc Karp & Woods, 2008; Lund & Veal, 2008). The items surrounding perceived importance of assessing

motor skills falls within the ‘support for motor skill assessment’ factor, thus representing a teacher’s perceived importance of assessment as a support mechanism for assessment. Originally seen as a barrier, the item referring to sharing results with students fell into the ‘support for motor skill assessment’ factor, thus indicating that teachers might perceive importance for this pedagogical skill. Historically, DinanThompson & Penney (2015) found that teachers often feel there is not enough time in physical education to share assessment results with students individually. Finally, a lack of time is consistently noted as a barrier to assessment in physical education (Lander et al., 2016, 2017; MacPhail& Halbert, 2010; Michael et al., 2016, Mintah, 2003). The belief that assessment is a good use of a physical education teacher’s time falls within the ‘support of motor skill assessment’ because it likely represents the perceived importance of motor skill assessment as a pedagogical skill.

‘Class Size and Time’

The second factor, ‘class size and time’, refers to the amount of time that physical education teachers have with students in physical education and the number of students in each class as either positive or negative influences over motor skill assessment. Too large of a class size and limited time in physical education are two of the most commonly stated barriers to conducting assessment in physical education. The class size survey items within this factor refer to class sizes being too large to adequately (quality) and frequently assess student’s motor skills. Referring to class time, there are two items purposely constructed to support each other, with one item stating ‘the time needed to formally assess motor skills negatively impacts the opportunity for my students to learn motor skills in class’ and the other asking if the teacher typically has enough time to

assess motor skills in class. As previously mentioned, the limited amount of class time (i.e., typically only a half hour; Imwold, Rider, & Johnson, 1982) typically occurring only once or twice per week present a challenge for how best to use instructional time (DinanThompson & Penney, 2015; Lander et al., 2016, 2017; MacPhail& Halbert, 2010; Michael et al., 2016, Mintah, 2003). Compounding the issue of limited class time, physical education teachers are also presented with large class sizes (Annerstedt & Larsson, 2010; Arslan et al., 2013; Michael et al., 2016), thus presenting a challenge for feasibly conducting assessment for all students in physical education in the given amount of time. The pattern of individual participant responses to these survey items that historically represent major barriers to assessment in physical education likely formed the factor 'class size and time' due to similar patterns of participant responses, either positively or negatively.

'Feasibility'

The third factor, 'feasibility', refers to constructs or elements that might influence the feasibility or likelihood of conducting motor skill assessment in physical education. Two items ask about a physical education teacher's perceived knowledge and training as influences over their ability to conduct assessment in physical education. Other items refer to the influence of an administrator's support (i.e., building principal) and having access to technology (if needed) as supports for conducting motor skill assessment in physical education. The final item within the 'feasibility' factor asks if physical education teachers believe that assessment takes places seamlessly in class, such that it does not disrupt instruction. Typically, physical education teachers perceive limited administrator and district support (Annerstedt & Larsson, 2010; Lander et al., 2015; 2017; Michael et

al., 2016; Slingerland et al., 2017), possibly including the use of and access to technology, as influences over their assessment behaviors. In line with limited class time and large class sizes, teachers typically recommend a need for authentic assessments that can feasibly be used in physical education (DinanThompson & Penney, 2015; Georgakis et al., 2015, Lander et al., 2017; Leirhaug & MacPhail, 2015). Finally, a major feasibility influence over whether physical education teachers assess in physical education come down to a lack of knowledge (DinanThompson & Penney, 2015; Lander et al., 2017; Leirhaug & Annerstedt, 2016a; Leirhaug & MacPhail, 2015), specifically in regard to motor skill assessment (Lander et al., 2015), and a lack of training or how to use assessment (Annerstedt & Larsson, 2010; Lander et al., 2017; Michael et al., 2016; Slingerland et al., 2017). The survey items falling within the 'feasibility' factor may positively or negatively influence a teacher's ability and decision to feasibly assess motor skills in physical education.

Limitations

This research is not without limitations. First, only 296 participants responded to individual (personal) demographic items including teacher age, sex, race, ethnicity, and number of years of teaching experience. Second, due to the nature of our small sample size (N = 445) and inability to split the data to be used for both an exploratory and confirmatory factor analysis, these data are without a confirmed factor structure. Ideally, exploratory and confirmatory factor analyses should be run on different sample sets (Froman, 2001). Since this study does not have a confirmed factor structure, the reliability of the findings should be interpreted with caution. Future research should consider obtaining a large enough sample size to run a confirmatory factor analysis on

survey items. Third, there are many items that are crossloading with other factors, thereby representing the need to run the EFA on another sample to see if the crossloading values change. Fourth, our data do not represent simple structure by having primary loading values with limited crossloading values. However, Bandalos (2018) does not recommend eliminating items with crossloading values until the items are tested on another sample. Therefore, future research should utilize this survey with another sample of elementary physical education teachers to obtain exploratory factor structure. Fifth, due to the nature of Hawaii's protection over teacher email addresses, teachers from this state were not included in our sample unless they responded to the survey via social media. Finally, due to the nature of using social media and other outlets (state representatives and SHAPE executives) to obtain a larger sample size, it is difficult to obtain an exact response rate for this survey.

Implications for Practice

The Elementary Physical Education Teacher Perception of Motor Skill Assessment Survey subscale has reliability (alpha = .82) and therefore can be used to interpret respondent results to the survey items. These findings should be interpreted with caution, however, being that the survey does not yet have a confirmed factor structure to declare that survey items are indeed measuring the latent construct 'perceptions'. With Cronbach's coefficient alpha in the acceptable range (>.80), this survey scale is recommended for use to measure elementary physical education teacher's perceptions of assessing students' motor skills. The Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey subscale can be used with inservice elementary physical education teachers to measure the perceived facilitators and barriers of

conducting motor skill assessment. Lander et al. (2015) recommends that physical education teacher perceptions likely motivate their assessment behaviors. Therefore, it is imperative to understand physical education teacher's perceptions of motor skill assessment prior to measuring their actual assessment behaviors. The findings from the perceptions survey subscale can inform teacher education programs and preservice professional development regarding the perceived facilitators and barriers toward conducting motor skill assessment at the elementary level. Teacher education programs and professional development programs can use the items housed within this survey to influence program goals and objectives. Furthermore, the survey can serve as a pre/post measure of teacher perceptions regarding teacher education or professional development programming. Understanding elementary physical education teacher perceptions of motor skill assessment will give researchers a glimpse of what may currently be happening across the United States of America.

Conclusion

The Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Survey Scale demonstrates content validity and internal consistency reliability using a sample of elementary physical education teachers across the United States of America. Therefore, this survey can be used to derive descriptive data from this population about their perceptions of motor skill assessment taking place at the elementary school level. This survey tool can be used, with caution due to its lack of confirmed factor structure, to measure teacher perception of motor skills across multiple samples to better generalize perception findings. Future research should investigate a

confirmed factor structure using a similar sample of elementary physical education teachers to solidify the psychometrics of this survey tool.

Table 4.1. Individual (Personal) Demographic Statistics for Survey Respondents

Individual (Personal) Demographic Items (participants, *n* = 296)		
Demographic Items & Categories	**Frequency**	**Percentage**
Sex		
Male	63	21.3%
Female	233	78.8%
Age		
20-25	16	5.4%
26-30	28	9.5%
31-35	54	18.2%
36-40	56	18.9%
41-45	36	12.2%
46-50	32	10.8%
51-55	34	11.5%
56-60	24	8.1%
61-65	16	5.4%
66-70	0	0%
71-75	0	0%
76-80	0	0%
80+	0	0%
Ethnicity		
Hispanic or Latino	12	4.1%
NOT Hispanic or Latino	274	92.6%
Unknown / Not Reported	10	3.4%
Race		
American Indian/Alaska Native	0	0%
Asian	6	2.0%
Native Hawaiian or Other Pacific Islander	1	0.3%
Black or African American	4	1.4%
White	280	94.6%
More Than One Race	4	1.4%
Unknown / Not Reported	1	0.3%
Years of Experience Teaching Physical Education		
1-5	58	19.6%
6-10	58	19.6%
11-15	56	18.9%
16-20	43	14.5%
21-25	28	9.5%
26-30	23	7.8%
30+	30	10.1%

Table 4.2. Demographic and School Context Statistics for Survey Respondents

Demographic and School Context Items (participants, *n* = 445)		
Demographic Items & Categories	**Frequency**	**Percentage**
Teaching Responsibility		
100% General PE	319	71.7%
75% General PE and 25% Adapted PE (APE)	100	22.5%
50% General PE and 50% APE	6	1.3%
25% General PE and 75% APE	2	.4%
100% APE	2	.4%
Other	16	3.6%
Size of Physical Education Department		
Only Teacher	193	43.3%
1 other	144	32.4%
2 others	22	4.9%
3 others	21	4.7%
4 others	9	2.0%
5 others	6	1.3%
6 others	3	.7%
7 others	3	.7%
8 others	5	1.1%
9 others	3	.7%
10 others	4	.9%
More than 10 others	32	7.2%
Time in Physical Education (Select all that apply)		
20-25 minutes	23	5.2%
26-30 minutes	99	22.2%
31-35 minutes	19	4.3%
36-40 minutes	79	17.8%
41-45 minutes	151	33.9%
46-50 minutes	66	14.8%
51-55 minutes	21	4.7%
56-60 minutes	20	4.5%
Over 60 minutes	5	1.1%
Other	1	.2%
Typical Class Size		
1-10 students	1	.2%
11-15 students	14	3.1%
16-20 students	56	12.6%
21-25 students	178	40%
26-30 students	105	23.6%
31-35 students	22	4.9%

36-40 students	8	1.8%
41-45 students	11	2.5%
46+ students	50	11.2%
Number of Times Teaching Each Class Per Week		
1 time	130	29.2%
2 times	190	42.7%
3 times	50	11.2%
4 times	17	3.8%
5 times	50	11.2%
Other	8	1.8%
Have a District Physical Education Supervisor		
Yes	128	28.8%
No	288	64.7%
I do not know	29	6.5%
Have a school policy for assessment in physical education		
Yes	122	27.4%
No	295	66.3%
I do not know	28	6.3%
Have a district policy for assessment in physical education		
Yes	167	37.5%
No	230	51.7%
I do not know	48	10.8%
Have a state policy for assessment in physical education		
Yes	157	35.3%
No	126	28.3%
I do not know	162	36.4%
Perceived support from building principal in respect to physical education		
Very unsupportive	22	4.9%
Unsupportive	14	3.1%
Mediocre	93	20.9%
Supportive	147	33%
Very supportive	169	38%
Type of School		
Public	396	89%
Private	22	4.9%
Charter	16	3.6%
Magnet	4	.9%
Montessori	1	.2%
Virtual/Online	2	.4%
Private Special Education	2	.4%
Other	2	.4%
Socioeconomic Status of School		
<25% free or reduced lunch	118	26.5%
25-50% free or reduced lunch	101	22.7%
>50% free or reduced lunch	197	44.3%
I do not know	29	6.5%

School Setting		
Urban	116	26.1%
Suburban	181	40.7%
Rural	141	31.7%
Other	7	1.6%
Estimated number of students in the school		
0-250	62	13.9%
251-500	181	40.7%
501-750	134	30.1%
751-1000	51	11.5%
1000+	17	3.8%
Title I School		
Yes	236	53%
No	176	39.6%
I do not know	33	7.4%

Table 4.3. Descriptive Statistics for Perception Items

Item	Mean	SD	Skewness	Skewness Standard Error	Kurtosis	Kurtosis Standard Error
imptchloc_item2	5.56	1.24	-3.18	.12	8.87	.23
impassessloc_item_81	4.37	1.45	-.79	.12	-.09	.23
imptchmanip_item_53	5.56	1.22	-3.15	.12	8.85	.23
impassessmanip_item83	4.33	1.44	-.78	.12	-.12	.23
perc_27	3.87	1.47	-.33	.12	-.89	.23
perc_27b	3.93	1.47	-.37	.12	-.87	.23
perc_67	4.09	1.42	-.41	.12	-.71	.23
perc_30	3.16	1.36	.07	.12	-.96	.23
perc_40	3.67	1.34	-.24	.12	-.74	.23
perc_68	5.24	.78	-1.06	.12	1.68	.23
perc_32	4.84	1.20	-1.15	.12	.95	.23
perc_36	3.77	1.36	-.32	.12	-.70	.23
perc_70	3.69	1.57	-.22	.12	-1.12	.23
perc_41	4.33	1.19	-.82	.12	.43	.23
perc_39	4.09	1.37	-.77	.12	-.18	.23
perc_72	3.89	1.53	-.39	.12	-.88	.23
perc_43	3.74	1.52	-.30	.12	-.94	.23

Table 4.4. Eigenvalues and Percent Variance Explained by Each Extracted Factor

Factor Number	**Eigenvalue (K1 Crit.)**	**% Variance Explained by Factor**
1	4.53	20.20 %
2	2.26	15.10 %
3	1.80	12.03 %
4	1.30	8.65 %
5	1.06	7.05 %
6	.90	5.96 %

Table 4.5. Factor Loadings and Communality Values for the Five Factor Structure

Items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Item Communalities
Perc 30	**.78**	.40	.08	-.24	.22	.55
Perc 36	**.73**	.20	.27	-.34	.29	.46
Perc 40	**.70**	.10	.19	-.48	.39	.51
Perc 67	**.62**	.49	.04	-.15	.15	.50
Perc 41	**.50**	-.01	.24	-.37	.45	.38
Perc 39	**.45**	.04	.32	-.22	.16	.24
Perc 70	.38	.09	.29	-.10	.11	.21
Perc 27	.34	**.95**	-.03	-.06	.02	.84
Perc 27b	.36	**.94**	-.01	-.06	.01	.84
Perc 68	.26	-.11	**.78**	-.08	.03	.41
Perc 32	.23	.03	**.75**	-.06	.06	.40
ImpAssess81	.36	.04	.09	**-.93**	.35	.77
ImpAssess83	.34	.04	.06	**-.93**	.33	.77
Perc 72	.31	-.01	.05	-.35	**.99**	.86
Perc 43	.28	-.02	.03	-.32	**.92**	.85

Table 4.6. Factor Correlation Matrix for Five Factor Structure

Factor	**1**	**2**	**3**	**4**	**5**
1	1.00				
2	.30	1.00			
3	.31	-.09	1.00		
4	-.37	.01	-.10	1.00	
5	.33	-.03	.07	-.38	1.00

Table 4.7. Factor Loadings and Communality Values for the Four Factor Structure

Item	Factor 1	Factor 2	Factor 3	Factor 4	Communalities
Perc 72	**.95**	.13	.15	-.37	.86
Perc 43	**.92**	.11	.11	-.34	.85
Perc 41	**.50**	.19	.43	-.42	.38
Perc 27b	.05	**.87**	.06	-.11	.84
Perc 27	.05	**.86**	.03	-.11	.84
Perc 67	.23	**.65**	.30	-.26	.50
Perc 30	.33	**.63**	.43	-.36	.55
Perc 68	.03	-.03	**.64**	-.07	.41
Perc 36	.37	.46	**.57**	-.45	.46
Perc 32	.05	.07	**.55**	-.05	.39
Perc 39	.21	.20	**.48**	-.27	.46
Perc 70	.15	.21	**.42**	-.15	.21
Impassessloc 81	.37	.18	.19	**-.91**	.77
Impassessmanip 83	.35	.18	.15	**-.91**	.77
Perc 40	.47	.37	.49	**-.56**	.51

Table 4.8. Factor Correlation Matrix for Four Factor Structure

Factor	**1**	**2**	**3**	**4**
1	1.00			
2	.20	1.00		
3	.24	.22	1.00	
4	-.44	-.26	-.27	1.00

Table 4.9. Factor Loading and Communality Values for the Three Factor Structure

Item	Factor 1	Factor 2	Factor 3	Communalities
Perc 72	**.78**	.13	.05	.86
Perc 43	**.74**	.08	.02	.85
Perc 40	**.63**	.37	.46	.51
ImpassessLoc 81	**.62**	.21	.24	.77
ImpassessManip 83	**.60**	.20	.21	.77
Perc 41	**.58**	.19	.38	.38
Perc 27b	.08	**.87**	.06	.84
Perc 27	.08	**.86**	.04	.84
Perc 67	.31	**.65**	.28	.50
Perc 30	.46	**.63**	.40	.55
Perc 68	.07	-.01	**.67**	.41
Perc 32	.07	.08	**.57**	.39
Perc 36	.51	.46	**.54**	.46
Perc 39	.30	.21	**.47**	.24
Perc 70	.19	.21	**.40**	.21

Table 4.10. Factor Correlation Matrix for Three Factor Structure

Factor	**1**	**2**	**3**
1	1.00		
2	.26	1.00	
3	.28	.23	1.00

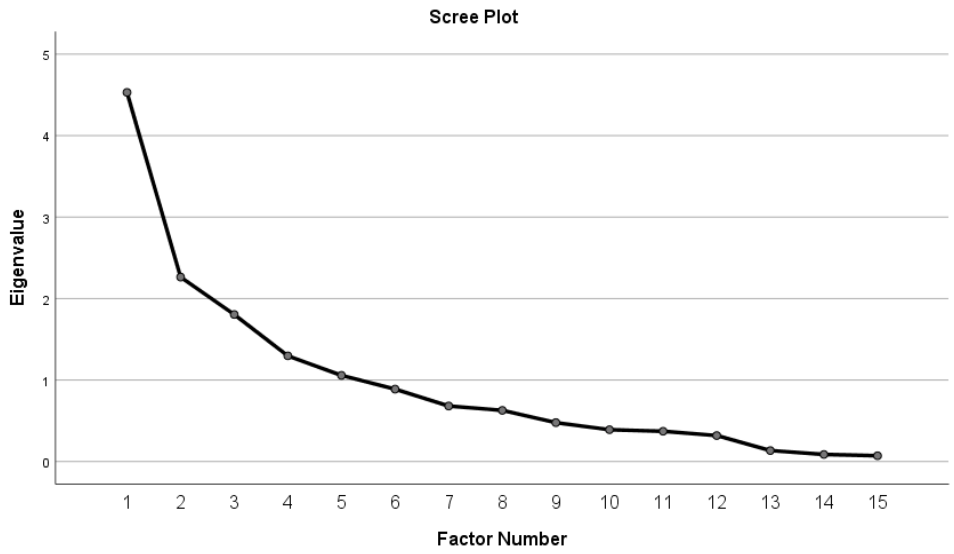

Figure 4.1. Scree Plot

CHAPTER 5

STUDY 3: DETERMINANTS OF ASSESSMENT BEHAVIOR USING THE ELEMENTARY PHYSICAL EDUCATION TEACHER PERCEPTION OF MOTOR SKILL ASSESSMENT SURVEY SUBSCALE

Introduction

A physical education teacher's perception of motor skill assessment likely influences the extent to which they actually assess (Lander et al., 2016; Leirhaug & MacPhail, 2015). Often the teacher's perceived barriers to assessment are the cause for assessment avoidance (Morgan & Hansen, 2008; Penney, 2012; Stiggins, 1997). Important is the examination of motor sill proficiency regarding the extent to which students are learning the basic motor skills that are needed to complete more complex motor skills in older grades in physical education (Castelli & van der Mars, 2018; Logan et al., 2015). Formal assessment is the documentation, recording, or collection of evidence of student progress toward learning (Lund & Tannehill, 2005) which is a critical step for physical education teachers to improve the teaching-learning process (van der Mars, Timken, & McNamee, 2018).

There is anecdotal evidence to suggest teacher's low perceptions of the utility of assessment in physical education (Castelli & van der Mars, 2018; Lopez-Pastor et al., 2013; van der Mars et al., 2018); however, limited research has investigated this issue (Emmanouilidou, Derri, Aggelousis, & Vassiliadou, 2012; Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, & Macdonald, 2013; Redelius & Hay, 2009; van der Mars et al.,

2018). Most notable, physical education teacher's perceptions of motor skill assessment at the elementary level are unknown (Fisher et al., in preparation; Imwold, Rider, & Johnson, 1982; Ni Chronin & Cosgrave, 2013).

Therefore, the purposes of this study were to: a) explore the determinants of motor skill assessment behavior, and b) predict the likelihood of motor skill assessment behavior using the perception survey factors. Our research questions were 'To what extent do assessment perceptions differ based upon whether teachers assess motor skills or not?' and 'Which motor skill assessment perception factors predict the likelihood for assessment of motor skills?'.

Methods

Design

This study is a descriptive analytic study with secondary data analysis using data from Study 2 to explore the determinants of assessment behavior using the Elementary Physical Education Teacher Perception of Motor Skill Assessment Survey subscale with a nationwide sample of inservice elementary physical education teachers.

Sample

Participants (N = 445) for this study were the same sample from Study 2. Participants included a voluntary national sample of elementary physical education teachers across the United States of America. Individual (personal) demographic item responses can be found in Table 4.1 while general demographic and school context item responses found in Table 4.2 (Tables presented after Study 2). A total of 296 participants (Female, n = 233, 78.7%) responded to personal demographic items with the majority of elementary physical education teachers between the ages of 31-35 (n = 54, 18.2%) and

ages 36-40 (*n* = 56, 18.9%) and white (*n* = 280, 94.6%), not Hispanic or Latino (*n* = 274, 92.6%). Of the 296 that responded to individual demographic items, the majority of elementary physical education teachers had between 1-5 (*n* = 58, 19.6%), 6-10 (*n* = 58, 19.6%), 11-15 (*n* = 56, 18.9%) years of teaching experience in physical education.

Instrumentation

The survey tool used for this study was the same Elementary Physical Education Teacher Perception Survey subscale used in Study 2. To measure elementary physical education teacher perceptions of motor skill assessment, we implemented the Elementary Physical Education Teacher Perception Survey subscale, only one of the survey scales demonstrating content validity from Study 1. The survey instrument demonstrated content validity using a modified Delphi method (see Study 1) and reliability (Cronbach coefficient alpha = .82; see Study 2). The survey is formatted electronically, collected, and managed by REDCap software (Harris et al., 2009; Harris et al., 2019) hosted by the University of South Carolina. The perception survey scale used in this study includes four behavior items, four importance items, and 13 perception items. The importance items were included within the perception scale the importance items measure perceived importance of teaching and assessing motor skills. Additionally, both scales use the same response options (e.g., six-point Likert scale from strongly disagree to strongly agree). See Appendix A for full perception survey.

Assessment behavior is measured using a dichotomous response scale (yes, no) with a focus on formal assessment behavior. An item example includes “Do you formally assess (written/documented) your students’ ability to perform locomotor skills (i.e., skip, jump, run) in your physical education classes?” and “Do you formally assess

(written/documented) your students' ability to perform manipulative skills (i.e., throw, catch, strike) in your physical education classes?" The responses to the assessment behavior items are meant to direct certain self-reported behaviors to different sections of the survey. For example, someone who responds that they do formally assess either locomotor or manipulative skills will respond to a select portion of survey items (including other assessment themes, like frequency of assessment), and the respondents who self-report that they do not assess locomotor or manipulative skills will respond to items solely related to perceptions of assessment.

Example items measuring the perceived importance of assessment of motor skills includes "Locomotor skills (e.g., run, jump, leap) should be taught in elementary school physical education" and "Locomotor skills (e.g., run, jump, leap) should be formally assessed (written/documented) in elementary physical education". Importance items fall on a six-point Likert-type scale (1 = strongly disagree, 3 = somewhat disagree, 4 = somewhat agree, 6 = strongly agree).

The item stem for perception related questions is: "To what extent do you agree that the following statements influence your use of formal motor skill assessment?" with subsequent item examples including "The time needed to formally assess motor skills negatively impacts the opportunity for my students to learn motor skills in class" or "I typically have enough time to formally assess motor skills". Perception items utilize a six-point Likert-type scale (1 = strongly disagree, 3 = somewhat disagree, 4 = somewhat agree, 6 = strongly agree).

The survey also includes demographic and school context questions, including teacher sex, years of teaching experience, location of school (i.e., rural vs. suburban),

Title I school, etc. Additionally, there is option to provide an email address for future correspondence to participate in a qualitative interview at a later date for individuals who wish to explain in greater detail their assessment perceptions. Due to our understanding and importance of teacher's time, the survey is constructed such that it can be completed in several increments, if needed, with the ability to save and finish the survey later. The survey should take no longer than 10 minutes to complete.

Perception Factors

Based on the proposed exploratory factor structure from Study 2, we separated the perception items (N = 15) into three factors. See Table 5.1 for a breakdown of item wording by factor. Consider that two items (item 2- importance to teach locomotor and item 53- importance to teach manipulative skills) were removed from the survey based on high skewness, kurtosis, and interitem correlations. The first factor, called "support for motor skill assessment" is made up of six items in support of motor skill assessment in elementary physical education. The second factor, called 'class size and time', includes four items referring to facilitators and barriers for assessment relating to a physical education class or the amount of time a teacher has to teach. The third factor, called 'feasibility' includes five items that support a teacher's ability to conduct motor skill assessment (e.g., knowledge, district support, and necessary technology).

Procedures

We acquired Institutional Review Board approval through the University of South Carolina prior to conducting this study. Prior to data analysis, the lead researcher checked the data to ensure all statistical assumptions are met. First, we ran the descriptive statistics for the perceptions scale by individual item and by factor (three-factor model

from Study 2). Next, we examined group differences by assessment behavior (assess or do not assess) for locomotor and manipulative skills separately by participant responses within the three-factor perception model. Finally, we evaluated which factors within the three factor perception model predicted self-reported locomotor skill assessment and manipulative skill assessment. Before running the independent samples *t*-tests for group differences and the binomial logistical regressions to determine predictions, we grouped survey items based on the proposed three factor structure from Study 2. Factors include support of motor skill assessment, class size and time, and feasibility. Participant responses to items within each of these factors were averaged, rather than summed, to improve interpretability of the findings for the reader.

Data Analysis

All statistical analyses were conducted using the Statistical Package for the Social Sciences (Version 26; SPSS Inc., Armonk, NY: IBM Corp.).

Descriptive Statistics

To better understand overall participant responses to individual items, we recorded the frequencies, means, and standard deviations for participant response by item. We interpreted these items in accordance with the six-point Likert scale for which they were set (1 = strongly disagree, 3 = somewhat disagree, 4 = somewhat agree, 6 = strongly agree). Given the descriptive nature of the study, higher perception ratings related to higher perceptions of assessment of motor skills. Note that Items 27, 27b, and 67 were reverse coded for negative wording.

We also recorded means and standard deviations for participant responses within the three factor perception model (Study 2). Participant responses are categorized into the

three perception factors ‘support for motor skill assessment’, ‘class size and time’, and ‘feasibility of assessment’.

Group Differences: Assessment Perception Rating versus Assessment Behavior

We wanted to determine group differences based on assessment behavior (assessors versus non-assessors) for locomotor skills and manipulative skills separately using participant responses within the three-factor perception model (support for motor skill assessment, class size and time, and feasibility).

Checking for Assumptions

Prior to analyzing our data for group differences, it is imperative that we check that there are equal variances (homogeneity) between our self-reported assessors and non-assessors for both locomotor and manipulative skills. We ran descriptive statistics to determine item variance by factor.

Next, we checked the data for normality using the Shapiro-Wilk Test. A significant *p*-value (<. 05) indicates non-normal distributed data while a non-significant *p*-value (> .05) represents normally distributed data. If our data is normally distributed, we can continue using the independent samples *t*-test. For non-normally distributed data, we will use the Mann-Whitney *U* test (Laerd Statistics, 2017).

Understanding we will use the Mann-Whitney *U* test for group differences, we needed to ensure our data met the assumptions required. Therefore, we needed to check that our dependent variable data (assessment perceptions) was ordinal or scale data and that our independent variable data (assessment behavior) was dichotomous. The Mann-Whitney *U* test also assumes observations are independent of one another. Finally, we needed to check that the independent variable distributions were similar to each other to

be able to compare medians. If the distributions are not similar, we will compare independent variable mean ranks. After checking the assumptions, we will run the Mann-Whitney *U* test for group differences and report the median or mean rank differences, the *U* statistic, standardized test statistic (*z* score), and the asymptotic significance level (2-sided test; Laerd Statistics, 2017).

Binomial Logistical Regression

We used binomial logistical regression to predict assessment behaviors for locomotor and manipulative skills separately using the three-factor perception model. Therefore, we wanted to be able to determine if responses within each of the perception factors would improve the odds of a teacher self-reporting whether they assess or do not assess motor skills (locomotor and manipulative skills). First, we checked to make sure our data met all the assumptions of binomial logistical regression. We used a Box Tidwell Test to ensure our data was linear in nature and the casewise list to check for outliers in the data (+- |2| standardized residuals). If normality and linearity within the data are present, we were able to continue with the binomial regression; however, if unmet, we needed to either transform the data or remove outliers (Laerd Statistics, 2017). After testing for assumptions, we ran binomial regressions to determine the extent to which the three factor model (Study 2) of perception items predicted assessment behavior (I assess versus I do not assess) for both locomotor and manipulative skills.

Results

Descriptive Statistics by Item

We calculated frequencies, means and standard deviations for participant responses to individual survey items which are out of a six-point scale (see Table 5.2).

Note that the interpretation of descriptive statistics utilizes original participant responses to items and not the items recoded for negative wording, as this would misrepresent participant responses. Participants rated highly on the importance to teach (M = 5.56, SD = 1.24; M = 5.56, SD = 1.22) and assess (M = 4.37, SD = 1.45; M = 4.33, SD = 1.44) both locomotor and manipulative skills. Individual perception items with the highest perceptions (i.e., strongly agree that these items influence assessment behavior) included items 67, 68, 32, 41, and 39. Generally, the highest rated perception items refer to time needed to assess (item 67: M = 4.09, SD = 1.42), motor skill assessment knowledge (item 68: M = 5.24, SD = .78), adequate training (item 32: M = 4.84, SD = 1.20), sharing assessment results with students (item 41: M = 4.33, SD = 1.93), and school district administrator support (item 39: M = 4.09, SD = 1.37). Each of the highest rated items, besides items 67 and 41, are housed within the 'feasibility' factor. Item 67 is a negatively worded item that asks teachers to what extent they believe that the time needed to formally assess motor skills negatively impacts a student's opportunity to learn motor skills in class. Item 41 refers to the sharing of assessment results with students. The lowest scoring item was item (M = 3.16, SD = 1.40) which asked teachers if they typically have enough time to formally assess motor skills.

Descriptive Statistics by Factor

We calculated the means and standard deviations for participant responses (N = 445) using the proposed perception factors from Study 2 which are out of a six-point response scale. The 'feasibility' factor had the highest average rating (M = 4.33, SD = .83), followed by the 'support for motor skill assessment' (M = 4.06, SD = 1.03) and 'class size and time' (M = 3.76, SD = .81; see Table 5.2) .

Group Differences for Perception Factors by Assessment Behavior

Descriptive Statistics for Assessment Behavior

Assessment behavior is made up of four items asking whether or not the elementary physical education teacher teaches locomotor and manipulative skills, and whether the teacher assesses locomotor and manipulative skills. Table 5.3 displays assessment behavior descriptive statistics. The majority of teachers declared they do teach locomotor skills (n = 440; 98.8%) and manipulative skills (n = 444; 99.8%). Interestingly, not all of the teachers who teach motor skills assess motor skills. Of the 98.8% of teachers who declare they teach locomotor skills, only 65.2% (n = 290) actually assess locomotor skills. Of the 99.8% of teachers who declare they teach manipulative skills, only 62.5% (n = 278) of the teachers actually assess manipulative skills.

Assumptions for Group Difference Testing

We checked for homogeneity of variance in the perception factor scores based on assessment behavior (assessors versus non-assessors) for locomotor and manipulative skills. Due to high variance in factor scores for factors 1 and 2 on both locomotor and manipulative skills (ranging from .71-1.30), likely due to large differences in sample size for assessors versus non-assessors, we violated the assumption of equal variance and therefore, considered using the non-parametric Mann-Whitney U test to determine group differences.

Assumptions for the Mann-Whitney U Test for Group Differences

In order to use the non-parametric Mann-Whitney U test for group differences, the dependent variable (assessment perceptions) must be continuous or ordinal. Our assessment perception scores are considered continuous data. Next, the independent

variable (assessment behavior) must be dichotomous (assess, do not assess). Finally, observations must be independent of each other. We have met all the assumptions for using the Mann-Whitney *U* test for group differences.

Additionally, we need to check our data to ensure the two independent variable distributions are similar using the population pyramid. If the two independent variable (assessors and non-assessors) distributions have the same shape, the Mann-Whitney *U* test can be used to compare group medians on the dependent variable (perception factors). If the two independent variable groups have different shapes, we must use the Mann-Whitney *U* test to compare mean ranks. Distributions for each of the assessment perception factor scores for assessors and non-assessors for locomotor and manipulative skills were not similar, as assessed by visual inspection (see Figures 3.1-3.6), and therefore, we will compare mean rank scores.

Mann-Whitney U Test for Group Differences

We used the Mann-Whitney *U* test to determine if there are group differences for each of the assessment perception factor scores (mean rank score) based on motor skill assessment behavior (whether the teacher assesses or not) for both locomotor and manipulative skills. See Table 5.4 and 5.5 for descriptive statistics regarding perception factor by type of motor skill assessment. Table 5.6 displays group differences in mean rank between assessment behavior for each assessment perception factor. In general, the non-assessors had statistically significantly lower mean rank scores (i.e., lower perceptions) than assessors by perception factor and by type of motor skill assessment (locomotor or manipulative).

‘Support for Motor Skill Assessment’. Assessment perception scores for the support factor for locomotor skill non-assessors (median = 3.33; mean rank = 148.17) were statistically significantly lower than locomotor skill assessors (median = 4.50; mean rank = 262.99; $U = 34073.00$, $z = 8.99$, $p < .001$). There is a mean rank difference of 114.82. Support for motor skill assessment perception scores were statistically significantly lower for manipulative skill non-assessors (median = 3.33; mean rank = 145.13) than for manipulative skill assessors (median = 4.50; mean rank = 269.78; ($U = 36217.50$, $z = 9.91$, $p < .001$). There is a mean rank difference of 124.65.

‘Class Size and Time’. Assessment perception scores for the class size and time factor for locomotor skill non-assessors (median = 2.25; mean rank = 156.87) was statistically significantly lower than locomotor skill assessors (median = 3.25; mean rank = 258.35; $U = 32725.50$, $z = 7.95$, $p < .001$). There is a mean rank difference of 101.48. Class size and time assessment perception scores for manipulative skill non-assessors (median = 2.25; mean rank = 169.92) were statistically significantly lower than for manipulative skill assessors (median = 3.25; mean rank = 254.89; $U = 32077.50$, $z = 6.76$, $p < .001$). There is a mean rank difference of 84.97.

‘Feasibility’. Feasibility assessment perception scores for locomotor skill non-assessors (median = 4.00; mean rank = 156.84) were statistically significantly lower than locomotor skill assessors (median = 4.60; mean rank = 258.36; $U = 32730.00$, $z = 7.96$, $p < .001$). There is a mean rank difference of 101.52. Feasibility assessment perception scores for manipulative skill non-assessors (median = 4.00; mean rank = 162.57) were also statistically significantly lower than manipulative skill assessors (median = 4.60;

mean rank = 259.30; $U = 33304.00$, $z = 7.71$, $p < .001$). There is a mean rank difference of 96.73.

Binomial Logistical Regression

Checking for Assumptions

We wanted to determine predictors of assessment behavior for locomotor and manipulative skills using the perception factors (support for motor skill assessment, class size and time, and feasibility). Prior to our analyses, we checked the linearity of our data using the Box Tindall Test (Table 5.7). A significant interaction between the factor and its logit would indicate nonlinearity as declared by a significant *p* value. Each factor interaction was not significantly related to it the dependent variable locomotor assessment behavior and manipulative skill assessment behavior. Our data was linear (i.e., the continuous independent variable is linearly related to the logit of the dependent variable), so we were able to proceed. Additionally, we checked the data for outliers +-|2| standard deviations away from the data (Laerd Statistics, 2017). We found 13 outliers and removed them from the data sample. Therefore, we ran the binomial logistical regressions on 432 responses.

Predicting Locomotor Skill Assessment Behavior

First, we ran a binomial logistical regression to predict assessment behavior for locomotor skills (i.e., to self-report the assessment of locomotor skills) using responses on the three factor perception model. The logistic regression model was statistically significant, $\chi 2(3) = 201.25$, $p < .001$. The model explained 51.7% (Nagelkerke R^2) of the variance in locomotor skill assessment behavior and correctly classified 79.9% of cases. Sensitivity was 88.85%, specificity was 62.07%, positive predictive value was 82.26%

and negative predictive value was 73.77%. All three of the predictor variables (perception factors) were statistically significant for predicting locomotor skill assessment behavior: support for motor skill assessment, class size and time, and feasibility (see Table 5.8). For every one-point value increase in support for motor skill assessment factor, participants had 3.10 times higher odds to self-report assessing locomotor skills. Likewise, for every one-point value increase in class size and time factor and the feasibility factor, participants were twice as likely (ExpB = 2.39, 2.42, respectively) to assess locomotor skills.

Predicting Manipulative Skill Assessment Behavior

Next, we ran a binomial logistical regression to predict assessment behavior for manipulative skills (i.e., to self-report the assessment of manipulative skills) using response to the three factor perception model. The logistic regression model was statistically significant, $\chi 2(3) = 180.41$, $p < .001$. The model explained 46.6% (Nagelkerke R^2) of the variance in manipulative skill assessment behavior and correctly classified 78.7% of cases. Sensitivity was 87.13%, specificity was 64.38%, positive predictive value was 80.61% and negative predictive value was 74.64%. All three of the predictor variables (perception factors) were statistically significant for predicting manipulative skill assessment behavior: support for motor skill assessment, class size and time, and feasibility (see Table 5.9). For every one-point value increase in support for motor skill assessment factor, participants were 3.34 times more likely to self-report assessing manipulative skills. Likewise, for every one-point value increase in class size and time factor and the feasibility factor, participants were nearly twice as likely (ExpB = 1.72, 2.02, respectively) to assess manipulative skills.

Discussion

The purpose of this study was to measure the significant determinants of the perception of motor skill assessment leading to assessment behavior and to predict assessment behavior using the motor skill perception survey with elementary physical education teachers across the United States of America.

Descriptive Statistics for Perceived Importance and Assessment Behavior

Elementary physical education teachers in this sample perceive locomotor and manipulative skills as important to teach and slightly less importance to assess. With almost the full sample of elementary physical education teachers declaring that they teach locomotor and manipulative skills, just over half declare they actually assess motor skills. There is a near 40% discrepancy between those who teach motor skills and those who assess motor skills. Our findings are relatively similar to those of Lander and colleagues (2015) who found in their Australian sample of middle school physical education teachers that most teachers taught, assessed, and perceived motor skill assessment as important within their motor skill instructional unit (Lander et al., 2015).

Descriptive Statistic by Item and Factor

Elementary physical education teachers were more likely to respond positively to perceived influences of motor skill assessment when they felt they had adequate knowledge and training of motor skill assessment to be able to perform motor skill assessment. Motor skill knowledge serving as a perceived influence over motor skill assessment perception was the highest rated item on the survey. Teachers also perceived school district administrator support as a big influence over their assessment practices. The aforementioned influences over assessment are informed by items that fall within the

'feasibility' factor of the perception survey, therefore, describing items that make assessment more feasible. Additionally, time persists as a barrier to perceptions of assessment (Kneer, 1986; Lander et al., 2015, 2016, 2017; Mintah, 2003; Morgan & Hansen, 2007, 2008). Within our sample, there was a large number of teachers (70.4%) who believed (somewhat agree to strongly agree) that the time needed to formally assess students' motor skills negatively impacted the students' ability to learn motor skills during class time. This finding is similar to that of Veal (1988), who discovered that with limited time in physical education, teachers were forced to make decisions about how best to use instructional time. Even more, teachers 'somewhat disagree to somewhat agree' that they have enough time to formally assess motor skills in physical education. The teachers' response to these items represents the negative perception that assessment takes away from learning time in physical education.

On average, items falling within the feasibility factor (i.e., knowledge, adequate training, administrator support, technology, and perceiving assessment as a seamless process) were rated the most highly of the three perception factors. With feasibility as a major influence over perceptions of motor skill assessment, it can be assumed that the elements of administrator support, motor skill knowledge, adequate motor skill assessment training, and access to technology have wide influence over whether or not an elementary physical education teacher perceives motor skill assessment in a positive or negative way. Similarly, elementary physical education teachers averaged 'somewhat agree' for the 'support for motor skill assessment' factor. Items within this factor relate to standardizing motor skill assessments across the state and nation, motor skill assessment as a good use of one's time, beliefs that motor skills (locomotor and manipulative) should

be taught in elementary physical education, and the belief of sharing assessment results with students. On average, elementary physical education teachers rated the support for motor skill assessment items as ‘somewhat agree’ as perceived influences over motor skill assessment. Surprisingly, teachers averaged between ‘somewhat disagree and somewhat agree’ that large class sizes and reduced amount of time in physical education serve as influences over their assessment practices. In summary, it seems that having adequate motor skill knowledge and training, in addition to administrator support, were the most influential determinants over one’s perception of motor skill assessment.

Group Differences: Perception Rating versus Assessment Behavior

We wanted to determine whether teachers who declare that they assess locomotor and manipulative skills or not respond differently on the perception survey. Locomotor and manipulative skill assessors perceived ‘support for motor skill assessment’ items (standardizing assessment, assessment importance, sharing results), ‘class time and size’ factors (large class sizes and not enough time), and ‘feasibility’ items (knowledge, training, administrator support, technology) as having a higher influence over their motor skill assessment practices than teachers who self-report not assessing locomotor and manipulative skills. Likely, due to the nature of the item wording and negatively worded items reverse coded to sound positive (for statistical analyses), teachers who do not assess motor skills also do not believe the same facilitators for assessment have influence over their assessment behaviors when compared to those who do assess motor skills. Similar to the findings of Lander et al. (2015), our sample of elementary physical education teachers also believed that motor skill assessment training influenced actual motor skill assessment behavior, with teachers claiming that adequate motor skill assessment training

positively influenced their assessment behaviors. Lander et al. (2015) also found that the physical education teachers in their sample who perceived importance over motor skill assessment were more likely to assess motor skills. Considering class size and time as a factor influencing motor skill assessment behavior, Mintah (2003) also found that large class sizes and lack of physical education instruction time impacted assessment behaviors.

Predicting Motor Skill Assessment Behaviors from Perception Factors

We used the perception survey factors to predict the likelihood of an elementary physical education teacher assessing both locomotor and manipulative skills in physical education. All of the perception survey factors were able to predict assessment behavior for both locomotor assessment and manipulative assessment. Therefore, an improvement in perception survey scores by just one response option (i.e., changing a perception from a ‘somewhat agree’ to an ‘agree’, in a positive direction) will make a teacher more likely to assess motor skills in elementary physical education. More specifically, a teacher who responds more favorably to the support for motor skill assessment factor items is three times more likely to assess locomotor and manipulative skills in physical education. Elementary physical education teachers are twice as likely to assess locomotor and manipulative skills in elementary physical education if they respond one response option more positively to feasibility factor items. Class size and time in physical education also predicted assessment behaviors, with a teacher twice as likely to assess if they respond more positively to survey items relating to influences of class size and time over assessment behavior. The question now remains, how do we get teachers to respond more positively to these perception items?

Limitations

One limitation to this research is the convenience survey sample. For example, one could assume that teachers who respond, or complete the survey, are teachers who have higher perceptions of assessment in physical education. Another limitation to this research is the exploratory nature of the proposed factor structure underlying the predictions of assessment behavior. As noted in Study 2, the sample size did not allow for a confirmatory factor analysis, thereby solidifying our ability to use the three factors (support for motor skill assessment, class size and time, and feasibility) a major predictor variables for motor skill assessment behavior. Since our sample produced reliable data using these items, we can say that the survey items were measuring the construct of perceptions of motor skill assessment behavior.

Implications for Practice

The elementary physical education teacher perception of motor skill assessment survey can be used to help determine factors relating to motor skill assessment behaviors at the elementary level. Elementary physical education teachers believe that it is important to teach motor skills in elementary physical education but less important to assess motor skills. This was evident in their self-reported assessment behaviors as well, with only just over half of the sample stating that they actually assess motor skills, compared to the rest who just teach the motor skills. Future research should investigate the elementary physical education teachers who do not assess motor skills to find out what barriers exist impeding their motor skill assessment practices. Elementary physical education teacher motor skill knowledge and adequate university training for motor skill assessment were two of the largest influences over assessment practices. Therefore,

universities, preservice teacher education programs, and professional development for inservice teachers should consider the implications of improving motor skill knowledge and training for assessing student motor skills could have over motor skill assessment practices at the elementary level. Another major influence over motor skill assessment practice is administrator (building principal) support for motor skill assessment occurring at the elementary level. Perhaps keeping the building principal and other administrator's privy to the assessment practices occurring in the gymnasium will help to improve overall physical education program support and advocacy among administration. Similar to the preservice teacher findings of Lund and Veal (2008), our sample of elementary physical education teachers might have distorted conceptions of the assessment process, such that it is wholly and completely distinct from the instructional process. Our sample of elementary physical education teachers perceived time as a major barrier to assessment, but most alarmingly, that the assessment of motor skills took time away from their student's ability to learn motor skills in physical education class. Assessment advocates would advise these physical education teachers that even during assessment of motor skills, students are continuing to gain practice trials, and therefore, with creative and deliberate assessment planning, are still learning. This misconception of assessment as a separate entity from instruction is something to be explored beginning with inservice teacher training. Additionally, teacher educators should consider providing and creating authentic assessments for motor skills such that teachers can use the motor skill assessment tools during physical education class, such that they do not feel assessment is a separate process from their instruction. Future research should investigate the use of authentic and 'easy to use' motor skill assessments for use in physical education classes,

regardless of short class time periods and large class sizes. Our survey was able to predict future motor skill assessment practices for each perception factor (support for motor skill assessment, class size and time, and feasibility) demonstrating that an improvement to perceptions of motor skill assessment influences can improve motor skill assessment behavior. Therefore, teacher education programming and professional development seminars should consider using the perception survey items to develop programing for the improvement of motor skill assessment behavior. For example, with perceived knowledge and adequate training as major influences over motor skill assessment behavior and class size and time as major barriers to the assessment for motor skills, programming should incorporate these elements into the program objectives and thus measure teacher perceptions before and after intervention. Additionally, future research should investigate other elements that influence the likelihood of elementary physical education teachers, especially the non-assessors, to assess (or avoid) motor skills. On a different note, future research should explore the frequency and types of motor skill assessment that are occurring in elementary physical education in addition to teacher's perceptions of motor skill assessment. Finally, this study should be replicated using a confirmed factor structure to ensure predictions hold true.

Conclusion

There is a large discrepancy between elementary physical education teachers who self-report teaching motor skills and those who self-report assessing motor skills. Future research should consider examining the differences between these two populations, as nearly 40% of the sample teaches motor skills but chooses not to assess locomotor and/or manipulative skills. Future research should consider measuring teacher motor skill

assessment frequency and the types of assessment used. Regarding the perception survey, the most important perception factors influencing assessment behavior were a teacher's perceived support for motor skill assessment and their perceived feasibility of assessment. With the ability to predict assessment behavior, it is important to consider ways to improve teacher perceptions of motor skill assessment in elementary physical education within the three factor domains (support, class size and time, and feasibility). Future research should look to discern the differences between teachers who do assess and do not assess motor skills and how to improve the likelihood of motor skill assessment in elementary physical education teachers.

Table 5.1. Perception Item Wording by Factor

Factor	Item	Item Wording
Factor 1 Support for Motor Skill Assessment	**Perc 72**	I believe that physical education should standardize motor skill assessments across the state so that each district in the state administers the same formal motor skill assessments.
	Perc 43	I believe that physical education should standardize motor skill assessments across the country so that each state administers the same formal motor skill assessments.
	Perc 40	I believe that formal motor skill assessment is a good use of my time as a physical education teacher
	Importance Assess. Loc. 81	Locomotor skills (e.g., run, jump, leap) should be formally assessed (written/documented) in elementary physical education.
	Importance Assess. Manip. 83	Manipulative skills (e.g., throw, catch, kick) should be formally assessed (written/documented) in elementary physical education.
	Perc 41	I believe that it is important to share formal motor skill assessment results with my students so they can use the information to improve their skills
Factor 2 Class Size and Time	**Perc 27b**	My class sizes are too large to formally assess motor skills as frequently (often) as I would like to
	Perc 27	My class sizes are too large to formally assess motor skills as well (quality) as I would like to
	Perc 67	The time needed to formally assess motor skills negatively impacts the opportunity for my students to learn motor skills in class
	Perc 30	I typically have enough time to formally assess motor skills
Factor 3 Feasibility	**Perc 68**	I feel knowledgeable enough about motor skills to formally assess my students
	Perc 32	I have received adequate training in my teacher certification program to be able to formally assess motor skills
	Perc 36	I believe that formal motor skill assessment takes place seamlessly in my class, such that it does not disrupt instruction
	Perc 39	My school district administrators (including building principal) support formal assessment of motor
	Perc 70	I believe that I have access to the technology that I would need to formally assess motor skills (if I choose to use technology)

Table 5.2. Descriptive Statistics for Perception Items and Perception Factors

Item	Frequencies						Mean (SD)	
	SD	**D**	**SWD**	**SWA**	**A**	**SA**	**Mean**	**SD**
imptchloc_item2	28 (6.3%)	N/A	N/A	5 (1.1%)	44 (9.9%)	368 (82.7 %)	5.56	1.24
impassessloc_item_81	30 (6.7%)	28 (6.3%)	35 (7.9%)	125 (28.1%)	110 (24.7 %)	117 (26.3 %)	4.37	1.45
imptchmanip_item_53	26 (5.8%)	1 (.2%)	N/A	8 (1.8%)	48 (10.8 %)	362 (81.3 %)	5.56	1.22
impassessmanip_item83	29 (6.5%)	31 (7.0%)	37 (8.3%)	120 (27.0%)	121 (27.2 %)	107 (24%)	4.33	1.44
perc_27 *negatively worded item	28 (6.3%)	80 (18.0%)	42 (9.4%)	129 (29.0%)	105 (23.6 %)	61 (13.7 %)	3.87	1.47
perc_27b *negatively worded item	27 (96.1%)	76 (17.1%)	43 (9.7%)	122 (27.4%)	111 (24.9 %)	66 (14.8 %)	3.93	1.47
perc_67 *negatively worded item	19 (4.3%)	60 (13.5%)	53 (11.9 %)	125 (28.1%)	105 (23.6 %)	83 (18.7 %)	4.09	1.42
perc_30	55 (12.4%)	104 (23.4%)	99 (22.2 %)	100 (22.5%)	75 (16.9 %)	12 (2.7%)	3.16	1.36
perc_40	28 (6.3%)	72 (16.2%)	82 (18.4 %)	131 (29.4%)	100 (22.5 %)	32 (7.2%)	3.67	1.34
perc_68	N/A	4 (.9%)	5 (1.1%)	55 (12.4%)	195 (43.8 %)	186 (41.8 %)	5.24	.78
perc_32	7 (1.6%)	22 (4.9%)	26 (5.8%)	79 (17.8%)	155 (34.8 %)	156 (35.1 %)	4.84	1.20
perc_36	28 (6.3%)	64 (14.4%)	76 (17.1 %)	129 (29.0%)	109 (24.5 %)	39 (8.8%)	3.77	1.36

perc_70	48 (10.8%)	76 (17.1%)	66 (14.8 %)	86 (19.3%)	113 (25.4 %)	56 (12.6 %)	3.69	1.57
perc_41	13 (2.9%)	28 (6.3%)	45 (10.1 %)	132 (29.7%)	167 (37.5 %)	60 (13.5 %)	4.33	1.19
perc_39	30 (6.7%)	40 (9.0%)	47 (10.6 %)	117 (26.3%)	163 (36.6 %)	48 (10.8 %)	4.09	1.37
perc_72	38 (8.5%)	68 (15.3%)	42 (9.4%)	122 (27.4%)	105 (23.6 %)	70 (15.7 %)	3.89	1.53
perc_43	44 (9.9%)	73 (16.4%)	47 (10.6 %)	127 (28.5%)	97 (21.8 %)	57 (12.8 %)	3.74	1.52
Factor 1: Support MSA	n/a	n/a	n/a	n/a	n/a	n/a	4.06	1.03
Factor 2: SizeTime	n/a	n/a	n/a	n/a	n/a	n/a	3.76	.81
Factor 3: Feasibility	n/a	n/a	n/a	n/a	n/a	n/a	4.33	.83

Note: SD = strongly disagree; D = disagree; SWD = somewhat disagree; SWA = somewhat agree; A = agree; SA = strongly agree

Table 5.3. Frequencies and Percentages for Assessment Behavior

	Yes **Frequency (Percent)**	**No** **Frequency (Percent)**
Teach Locomotor Skills	440 (98.9%)	5 (1.1%)
Teach Manipulative Skills	444 (99.8%)	1 (.2%)
Assess Locomotor Skills	290 (65.2 %)	155 (34.8%)
Assess Manipulative Skills	278 (62.5%)	167 (37.5%)

Table 5.4. Descriptive Statistics for Perception Factors and Locomotor Assessment Behavior

Factor	Assess Locomotor	N	Mean	SD	Std. Error Mean
F1of3_Support	No	155	3.42	1.06	.09
	Yes	290	4.40	.84	.05
F2of3_SizeTime	No	155	3.94	.82	.07
	Yes	290	3.67	.80	.05
F3of3_Feasibility	No	155	3.90	.80	.06
	Yes	290	4.56	.75	.04

Table 5.5. Descriptive Statistics for Perception Factors and Manipulative Assessment Behavior

	Assess Manipulative	**N**	**Mean**	**SD**	**Std. Error Mean**
F1of3_Support	No	167	3.41	1.05	.08
	Yes	278	4.45	.81	.05
F2of3_SizeTime	No	167	3.86	.83	.06
	Yes	278	3.70	.80	.05
F3of3_Feasibility	No	167	3.95	.83	.06
	Yes	278	4.56	.73	.04

Table 5.6. Mann-Whitney U Group Differences (Mean Rank) for Perception Factors by Assessment Behavior

Locomotor Assessment Behavior Mean Rank				
Factor	**Assessors**	**Non-Assessors**	***U (p value)***	**Z**
Perception Factor 1 Support for Motor Skill Assessment	262.99	148.17	34073.00**	8.99
Perception Factor 2 Class Size and Time	258.35	156.87	32725.50**	7.95
Perception Factor 3 Feasibility	258.36	156.84	32730.00**	7.96
Manipulative Skill Assessment Behavior Mean Rank				
Factor	**Assessors**	**Non-Assessors**		
Perception Factor 1 Support for Motor Skill Assessment	269.78	145.13	36217.50**	9.91
Perception Factor 2 Class Size and Time	254.89	169.92	32077.50**	6.76
Perception Factor 3 Feasibility	259.30	162.57	33304.00**	7.71
$p < .05$ denoted by *				
$p < .001$ denoted by **				

Table 5.7. Box Tidwell Procedure for Linearity

Assessment of Locomotor Skills								
	B	**S.E.**	**Wald**	**Df**	**Sig.**	**Exp(B)**	**Lower CI for EXP (B)**	**Upper CI for EXP (B)**
F1of3_Support by lnF_1of3_Support	-.63	.80	.62	1	.43	.53	.11	2.56
F2of3_TimeSize.revcode by lnF_2of3_SizeTime.revcode	-.02	.55	.00	1	.97	.98	.34	2.85
F3of3_Feasibility by lnF_3of3_Feasibility	-.10	1.15	.01	1	.93	.90	.10	8.58
Constant	-9.97	5.02	3.95	1	.05	.000		
Assessment of Manipulative Skills								
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower CI for EXP(B)**	**Upper CI for EXP(B)**
F1of3_Support by lnF_1of3_Support	-.50	.87	.33	1	.57	.61	.11	3.36
F2of3_TimeSize.revcode by lnF_2of3_SizeTime.revcode	-.80	.51	2.48	1	.12	.45	.17	1.21
F3of3_Feasibility by lnF_3of3_Feasibility	-.65	1.13	.33	1	.57	.52	.06	4.75
Constant	-13.74	5.24	6.88	1	.01	.00		

a. Variable(s) entered on step 1: F1of3_Support, F2of3_TimeSize.revcode, F3of3_Feasibility, F1of3_Support * lnF_1of3_Support, F2of3_TimeSize.revcode * lnF_2of3_SizeTime.revcode, F3of3_Feasibility * lnF_3of3_Feasibility.

Table 5.8. Binomial Logistic Regression Predicting Likelihood of Locomotor Assessment Behavior based on Three Perception Factors

							95% C.I. for EXP(B)	
Factors	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
F1of3_Support	1.13	.16	50.31	1	.00*	3.10	2.27	4.23
F2of3_TimeSize	.87	.15	35.60	1	.00*	2.39	1.80	3.19
F3of3_Feasibility	.88	.18	22.97	1	.00*	2.42	1.69	3.47
Constant	-9.91	1.08	84.89	1	.00	.00		

Variable(s) entered on step 1: F1of3_Support, F2of3_TimeSize.revcode, F3of3_Feasibility.

Table 5.9. Binomial Logistic Regression Predicting Likelihood of Manipulative Assessment Behavior based on Three Perception Factors

							95% C.I. for EXP(B)	
Factors	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
F1of3_Support	1.21	.15	62.31	1	.00*	3.34	2.48	4.51
F2of3_TimeSize	.54	.13	18.15	1	.00*	1.72	1.34	2.20
F3of3_Feasibility	.70	.17	16.73	1	.00*	2.02	1.44	2.82
Constant	-8.78	.97	82.21	1	.00	.00		

Variable(s) entered on step 1: F1of3_Support, F2of3_TimeSize.revcode, F3of3_Feasibility.

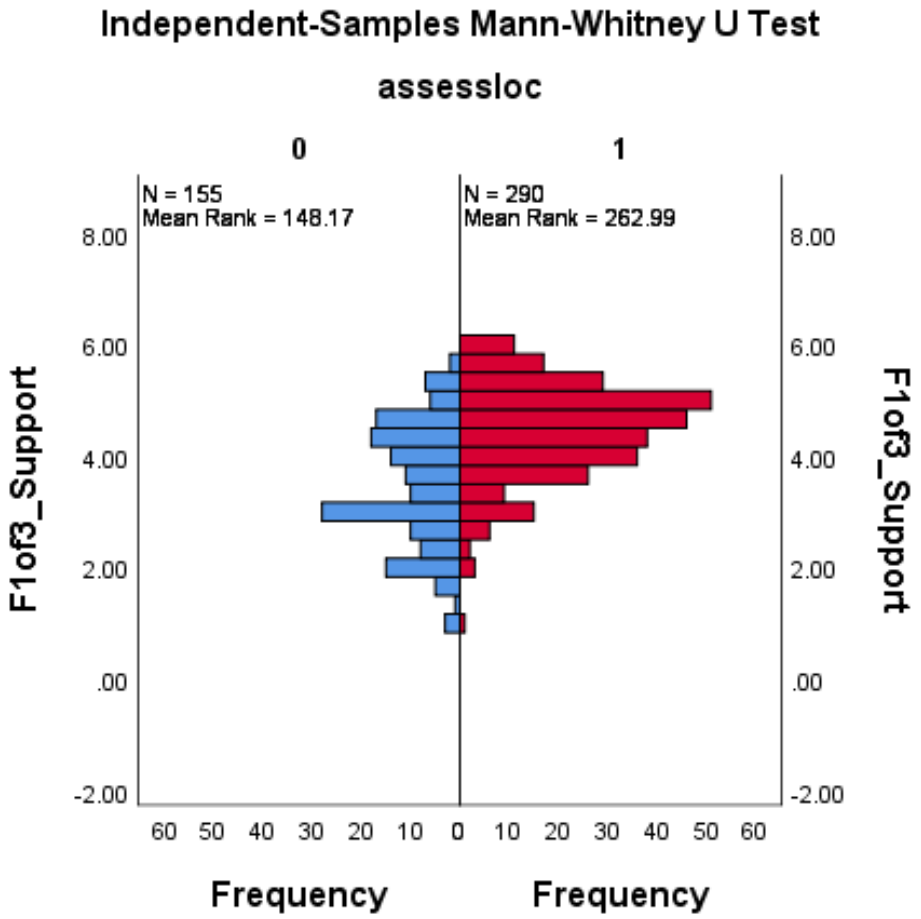

Figure 5.1. Independent Samples Mann-Whitney U Test for Perception Factor 1 and Locomotor Skill Assessment Behavior

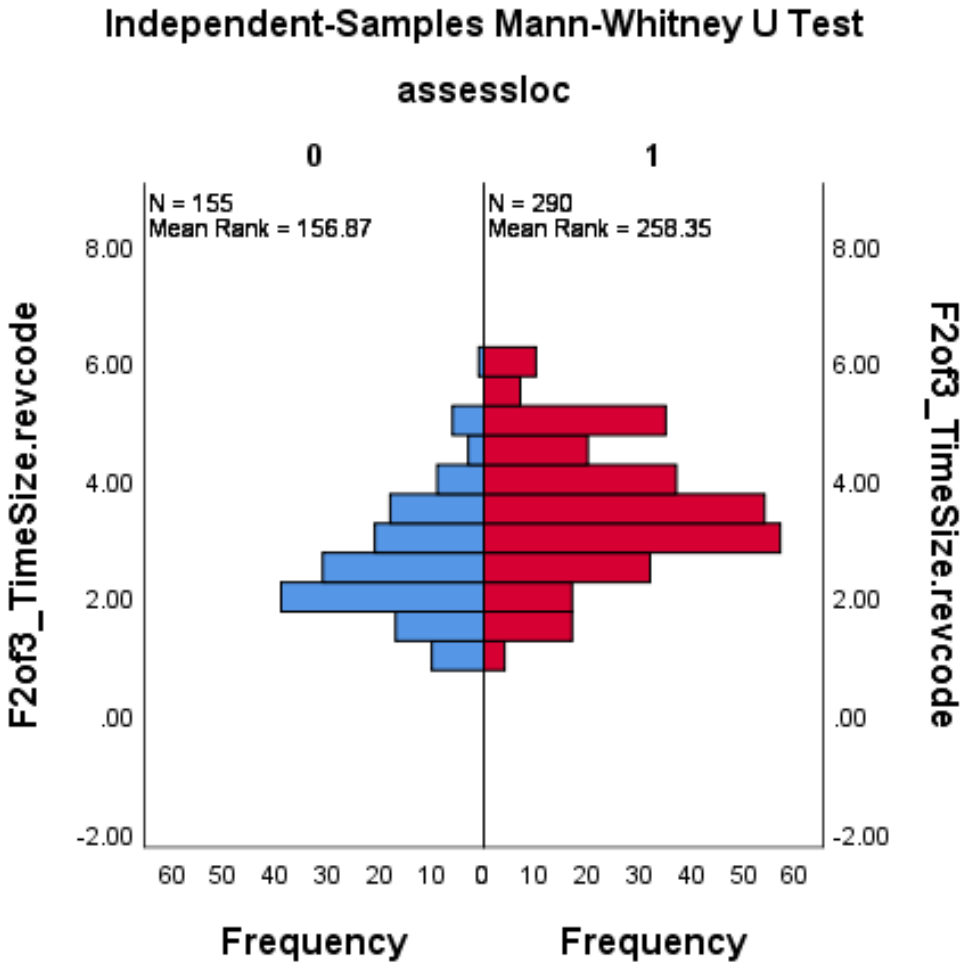

Figure 5.2. Independent Samples Mann-Whitney U Test for Perception Factor 2 and Locomotor Skill Assessment Behavior

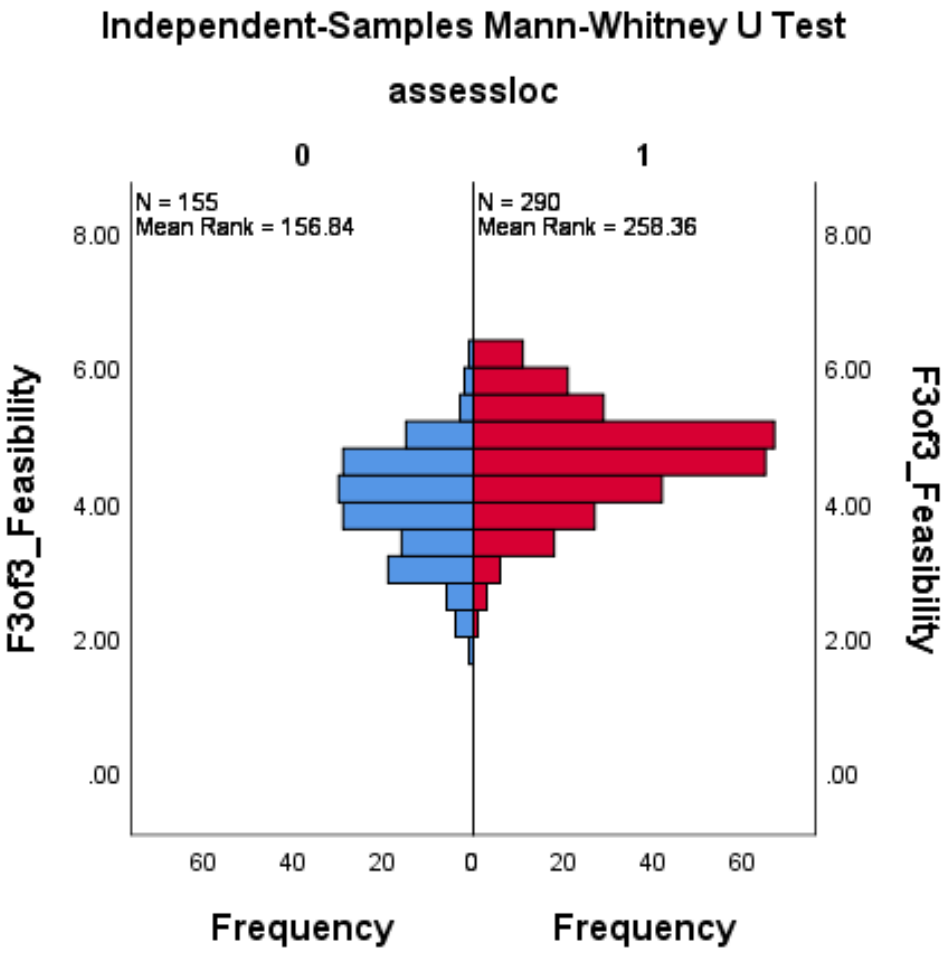

Figure 5.3. Independent Samples Mann-Whitney U Test for Perception Factor 3 and Locomotor Skill Assessment Behavior

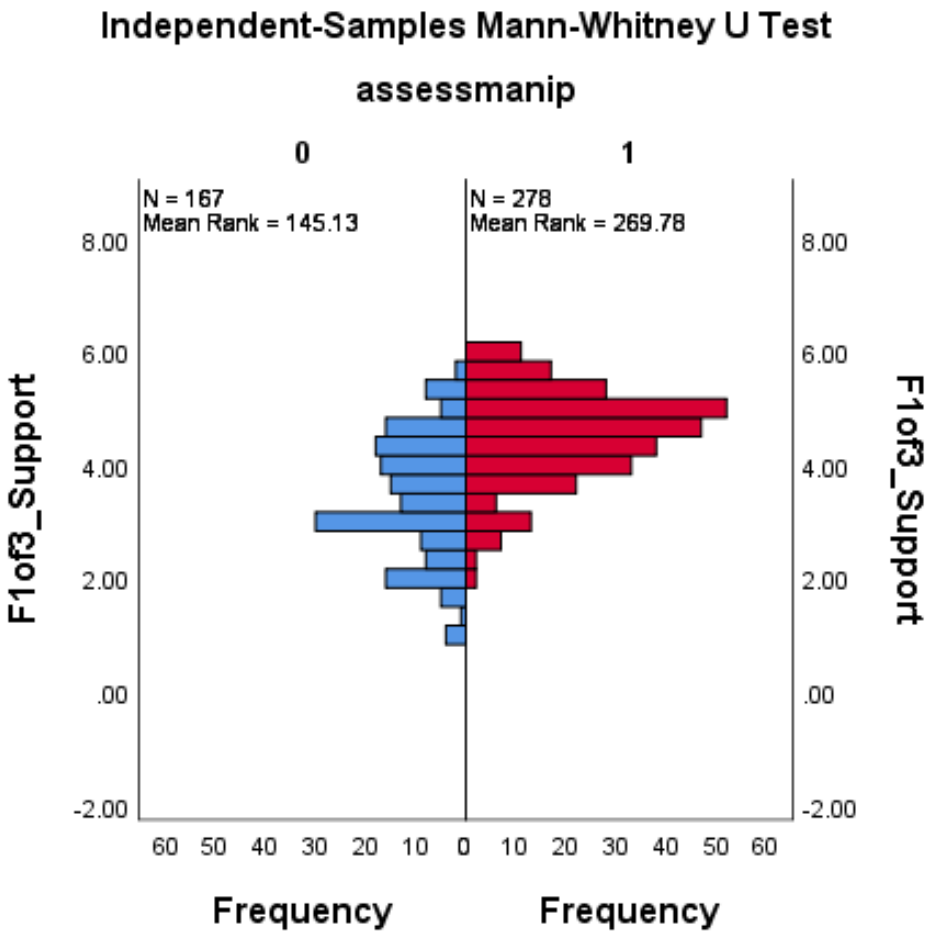

Figure 5.4. Independent Samples Mann-Whitney U Test for Perception Factor 1 and Manipulative Skill Assessment Behavior

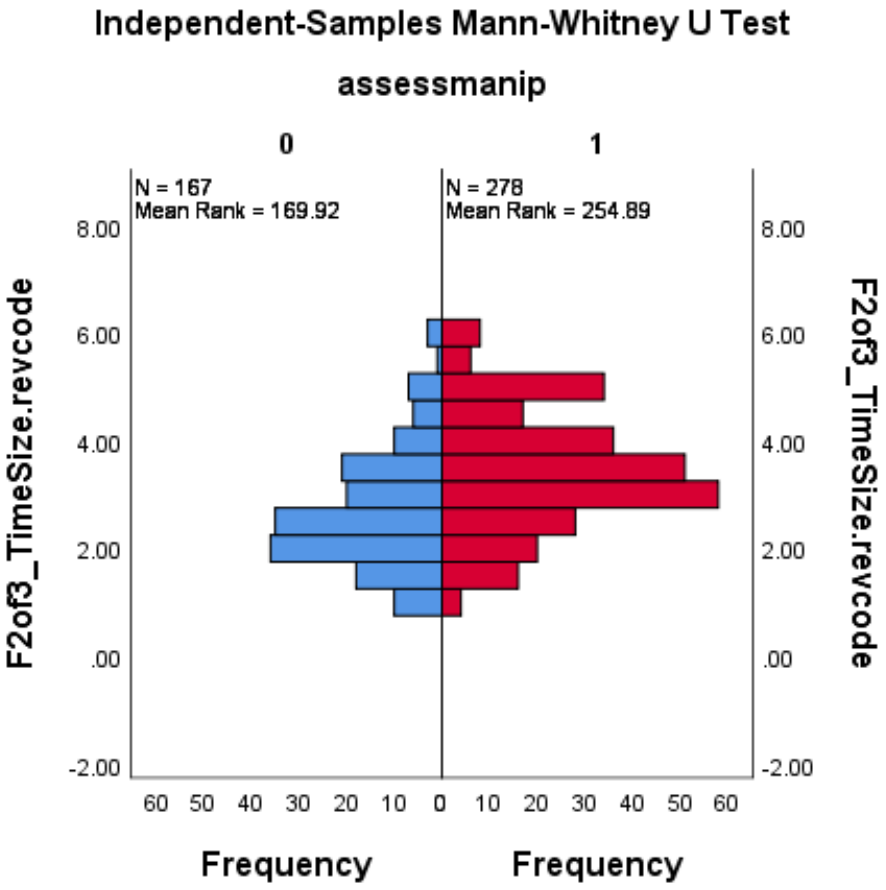

Figure 5.5. Independent Samples Mann-Whitney U Test for Perception Factor 2 and Manipulative Skill Assessment Behavior

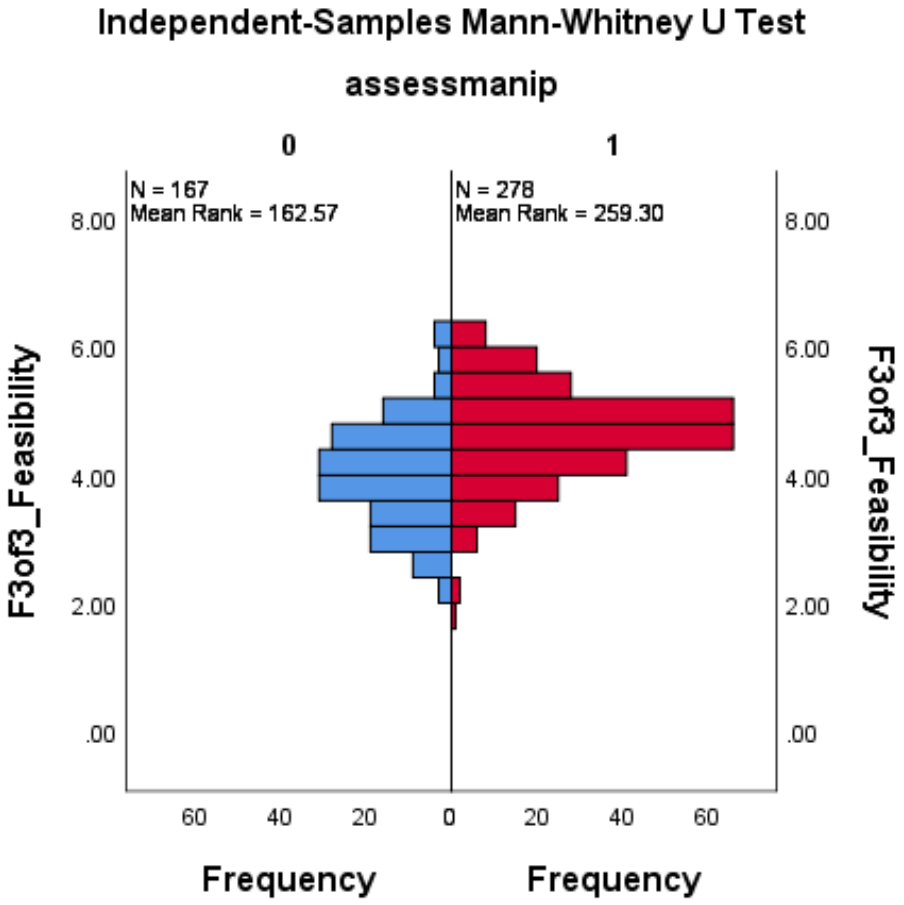

Figure 5.6. Independent Samples Mann-Whitney U Test for Perception Factor 3 and Manipulative Skill Assessment Behavior

CHAPTER 6

DISCUSSION, FUTURE RESEACH CONSIDERATIONS, & CONCLUSION

Discussion

The three studies contained within this dissertation contribute to physical education assessment literature as a whole. Physical education assessment research is uncommon (van der Mars et al., 2018), therefore, every quality contribution has the ability to advance the field. Physical education assessment literature at the elementary level is commonly neglected (Imwold, Rider, & Johnson, 1982; Ni Chronin & Cosgrave, 2013) and wholly neglected regarding the assessment of motor skills (see Fisher et al., in preparation). Therefore, we decided to use the extant physical education assessment literature to support survey item development to measure elementary physical education teacher motor skill assessment behaviors, specifically related to motor skill assessment perception.

Overall these studies addressed gaps in physical education assessment literature by a) measuring elementary teacher perceptions of assessment which has never been done before in the United States of America, b) measuring teacher perception of motor skill assessment which has also never been done before in the United States of America, and c) utilized a national sample of physical education teachers which has never been done in assessment literature in the United States of America. Most importantly, elementary physical education teachers, as the primary teachers of motor skills to children, have

never been asked their perceptions of formally measuring student success in motor skills through the use of assessment. Lander et al. (2015) believes that a teacher's perception of assessment will likely guide their decisions for assessment behavior. Therefore, we began our journey toward understanding motor skill assessment behavior at the elementary level by measuring teacher perceptions of this process.

Specifically, Study 1 served to as survey item development and content validation of survey items and subscales used to measure elementary physical education teacher motor skill assessment behaviors. Physical education assessment content experts and current inservice physical education teachers participated in the modified Delphi method to obtain consensus on survey item content validity. The "Elementary Physical Education Teacher Motor Skill Assessment Behavior Survey" can be broken into several subscales, including a motor skill perception subscale, motor skill frequency and type of assessment subscale, improved likelihood of motor skill assessment subscale, motor skill assessment environment subscale, motor skill assessment resources subscale, motor skill assessment learning opportunities subscale, and an Every Student Succeeds Act (E.S.S.A.) subscale.

Study 2 consisted of the dissemination and data collection using one of the survey subscales from Study 1, the "Elementary Physical Education Teacher Perceptions of Motor Skill Assessment Subscale". The perception survey subscale demonstrated content validity during Study 1 and exploratory factor structure and internal consistency during Study 2. We measured the Cronbach's coefficient alpha (.82) demonstrating reliability across the perception survey item subscale. Using an exploratory factor analysis, we explored the psychometric properties of this subscale using a sample of 445 current elementary physical education teachers across the United States of America. Due to

limitations in survey size, we were only able to run an exploratory factor analysis and obtained a sufficient three-factor structure. Factors that make up the perception survey subscale include: ‘support for motor skill assessment’, ‘class size and time’, and ‘feasibility’. Therefore, with acceptable internal consistency reliability, but only an exploratory factor structure, we can interpret findings from survey responses with caution. Future research should consider running a confirmatory factor analysis on these survey items with a similar sample of elementary physical education teachers.

Study 3 was a secondary data analysis of perception survey responses from the same sample in Study 2. We were able to ascertain descriptive information regarding elementary physical education teacher perceptions of motor skill assessment and make predictions for how to improve motor skill assessment behavior within this sample. Results from these analyses show that almost all elementary physical education teachers perceive that teaching motor skills (locomotor and manipulative skills) are important in elementary physical education, but a large portion of them (40%) declare that they do not assess these skills. Teachers who self-report that they do assess motor skills in elementary physical education typically had higher perception of motor skill assessment. Those who self-report that they do not assess motor skills in elementary physical education typically had lower perceptions of motor skill assessment. Perception items that received the highest perception rating, thus indicating that these elements are major influences over elementary physical education teacher motor skill assessment behaviors, include: adequate knowledge of motor skills to be able to assess, adequate university training for motor skill assessment to be able to assess, the importance of sharing motor skill assessment results with students, and administrator (building principal) support for motor

skill assessment. Therefore, these items can serve as major factors to improve motor skill assessment behavior in elementary physical education teachers moving forward. The factors within the perception survey subscale were able to predict the likelihood of improving motor skill assessment behaviors of elementary physical education teachers. Improving survey responses to perception items as a whole would make an individual between two and three times more likely to improve their locomotor and manipulative assessment behavior as a whole. Specifically, improving survey responses to items within the 'support for motor skill assessment' would make an individual three times more likely to assess locomotor and manipulative skills. Also, improving perception survey responses to items within the 'class size and time' and 'feasibility' factors would make an individual two times more likely to improve locomotor and manipulative skill assessment behaviors. Therefore, with the ability to predict motor skill assessment behaviors based on perception survey scores, these survey items can be used during intervention strategies (i.e., teacher education or inservice professional development) to measure changes in teacher perceptions over time. Additionally, these survey items can serve to support program goals and objectives in teacher education and professional development programming relating to motor skill assessment training.

Future Research Considerations

First, future studies should evaluate the confirmed factor structure of the proposed three-factor structure found within our exploratory factor analysis. Due to limitations in sample size, we were unable to measure the confirmed factory structure. Next, we only used one of the many assessment behavior subscales (perception subscale) with our sample of elementary physical education teachers. Each of the survey subscales from

Study 1 have content validity, and therefore should be used to measure different elements of assessment behavior across this sample. Based on our findings in Study 3, there seem to be many factors within motor skill assessment perception that influence the actual locomotor and manipulative skill assessment behaviors of elementary physical education teachers. These perceptions of motor skill assessment elements should be explored in greater detail, perhaps during motor skill assessment preservice teacher education programing or inservice professional development programming, looking for ways to improve the likelihood of elementary physical education teachers assessing motor skills.

Conclusion

This dissertation represented the first studies to measure assessment practices and perceptions of motor skill using elementary physical education teachers across the United States of America. Up until this point, elementary physical education teacher perceptions of assessment have not been measured, especially regarding the assessment of motor skills. Our nationwide sample of elementary physical education teachers provides the ability to generalize across the country the perceptions of motor skill assessment occurring in most elementary physical education settings. The information gleaned from the perceptions survey subscale can be used to improve motor skill assessment behaviors for preservice physical education teachers and current inservice physical education teachers. Most notably, the findings from the perception survey subscale demonstrate the major influences over actually assessing motor skills at the elementary physical education level are adequate knowledge about motor skills to be able to assess, adequate teacher education (university) training to be able to assess motor skills, a perceived importance over sharing motor skill assessment results with students, and the continued influence of

administrator (building principal) support for motor skill assessment. Future research should consider the implications of these perceived influences over motor skill assessment behavior to determine the actual motor skill assessment behaviors occurring at the elementary level with physical education teachers in the United States of America.

REFERENCES

American Education Research Association. (2014). Standards for educational and psychological testing. Washington, D.C.: American Education Research Association.

Annerstedt, C. & Larsson, S. (2010). 'I have my own picture of what the demands are…': Grading in Swedish PEH- problems of validity, comparability and fairness. *European Physical Education Review*,16(2), 97-115.

Arias-Estero, J. & Castejon, F. (2014). Using instruments for tactical assessment in physical education and extra-curricular sports. *European Physical Education Review*, 20(4), 525-535.

Avella, J. R. (2016). Delphi panels: Research design, procedures, advantages, and challenges. *International Journal of Doctoral Studies*, 11, 305-321.

Bailey, R. (2001). Teaching physical education: A handbook for primary & secondary school teachers. London, England: Psychology Press.

Bandalos, D.L. (2018). Measurement theory and applications for the social sciences. New York, New York: Guildford Press.

Bayless, J. (1978). Conflicts and confusion over evaluation. *Journal of Health, Physical Education, Recreation, and Dance*, 62, 21-23.

Benson, J.J. & Clark, F. (1982). A guide for instrument development and validation. *The American Journal of Occupational Therapy*.

Black, P. & Wiliam, D. (1998). Assessment in education: Principles, policy and practice. London, England: Routledge.

Black, P. & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Kappan Classic*, 92(1), 81-90.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. 2003. Assessment for learning: Putting it into practice, Maidenhead: Open University Press.

Bollen, K.A. (1989). Structural equations with latent variables. New York: Wiley.

Borghouts, L., Slingerland, M., & Haerens, L. (2017). Assessment quality and practices in PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473-489.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, Massachusetts: Harvard University Press.

Council for the Accreditation of Educator Preparation (CAEP; 2015). CAEP accreditation standards.

Carre, F.A., Mosher, R., Schutz, R., Thompson, R., Farenholtz, E., & Bullis, L. (1983). British Columbia assessment of physical education, general report. Province of British Columbia: The Ministries of Education.

Chen, W. (2005). Examination of curricula, teaching practices, and assessment through national standards. *Physical Education and Sport Pedagogy*, 10(2), 159-180.

Chen, W., Hammond-Bennett, A., & Hypnar, A. (2017). Examination of motor skill competency in students: Evidence-based physical education curriculum. *BMC Public Health*, 17(1), 1-8.

Coker, G.E. (1972). A survey of senior high school physical education programs for boys in selected Louisiana public schools. *Dissertation Abstracts International*, 33.

Database of Accredited Postsecondary Institutions and Programs (DAPIP; n.d.). The role of U.S. department of education in postsecondary accreditation. Retrieved from https://ope.ed.gov/dapip/#/home

Davis, L.L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194-197.

Desrosiers, P., Genet-Volet, Y., & Godbout, P. (1997). Teacher's assessment practices viewed through the instruments used in physical education classes. *Journal of Teaching in Physical Education*¸16, 211-228.

DinanThompson, M., & Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review*, 21(4), 485-503.

DiStefano, C. & Dombrowski, S.C. (2006). Investigating the theoretical structure of the Stanford-Binet fifth edition. *Journal of Psychoeducational Assessment*, 24(2), 123-136.

Doolittle, S. (1996). Practical assessment for physical education teachers. *Journal of Physical Education, Recreation, & Dance,*67(8), 35-37.

Emmanouilidou, K., Derri, V., Aggelousis, N., & Vassiliadou, O. (2012). Development and evaluation of a questionnaire to assess physical educators' knowledge of student assessment. *The Physical Educator.*

Frapwell, A. (2010). Assessment in physical education – fit for purpose? We've got an APP for that!. *Physical Education Matters*, 5(3), 13–8.

Froman, R.D. (2001). Elements to consider in planning the use of factor analysis. *Southern Online Journal of Nursing Research,* 5(2), 1-22.

Fox, K. (1959). The reliability and validity of selected physical fitness tests for high school girls. *Research Quarterly. American Association for health, Physical Education and Recreation, 30(4), 430-437.*

Georgakis, S., Wilson, R., & Evans, J. (2015). Authentic assessment in physical education: A case study of game sense pedagogy. *The Physical Educator*, 72, 67-86.

Goluchowics, K. & Blind, K. (2011). Identification of future fields of standardization: An explorative application of the Delphi methodology. Technological Forecasting & Social Change, 78, 1526-1541.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J.G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology

and workflow process for providing translational research informatics support. *J Biomed Inform, 42(2), 377-81*.

Harris, P.A., Taylor, R., Minor, B.L., Elliot, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S.N. (2019). REDCap Consortium, The REDCap consortium: Building an international community of software partners. *J Biomed Inform, 95, 1-10.*

Harrow, A. (1972) A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives. New York: David McKay.

Hay, P. J. (2006). "Assessment for learning in physical education". In The handbook of physical education, Edited by: Kirk, D., Macdonald, D. and O'Sullivan, M. 312–25. London: Sage

Hay, P. & Macdonald, D. (2008). Evidence for the social construction of ability. *Physical Education, Sport, Education and Society*, 15(1), 1-18.

Hay, P. & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review*, 15(3), 389-405.

Haynes, J. & Miller, J. (2015). Preparing pre-service primary school teachers to assess fundamental motor skills: Two skills and two approaches. *Physical Education and Sport Pedagogy*, 20(4), 397-408.

Hensley, L. (1997). Alternative assessment for physical education. *Journal of Physical Education, Recreation, and Dance*, 68(7), 19–24.

Hensley, L.D., Lambert, L.T., Baumgartner, T.A., & Stillwell, J.L. (1987). Is evaluation worth the effort? *Journal of Physical Education, Recreation, & Dance,* 58(6), 59-62.

Henson, R.K. & Roberts, J.K. (2006). Use of exploratory factor analysis in published research. *Journal of Psychological Measurement,* 66(3), 393-416.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika,* 30, 179-185.

Imwold, C., Rider, R., & Johnson, D. (1982). The use of evaluation in public school physical education programs. *Journal of Teaching in Physical Education,* 2(1), 13-18.

James, A.R., Griffin, L., & Dodds, P. (2009). Perceptions of middle school assessment: An ecological view. *Physical Education and Sport Pedagogy*, 14(3), 323-334.

Johnson, R.L. & Morgan, G.B. (2016). *Survey Scales: A Guide to Development, Analysis, and Reporting.* Guilford Publications: New York, NY.

Kaiser H.F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.

Kline, R.B. (2011). *Principles and practice of structural equation modeling. Third edition*. New York: Guilford Press.

Kneer, M.E. (1986). Description of physical education instructional theory/practice gap in selected secondary schools. *Journal of Teaching in Physical Education,* 5, 91-106.

Laerd Statistics. (2017). Binomial logistic regression using SPSS Statistics. Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Laerd Statistics. (2017). Mann-Whitney U test using SPSS Statistics. Statistical tutorials and software guides. Retrieved from https://statistics.laerd.com/

Lander, N.J., Barnett, L.M., Brown, H. & Telford, A. (2015). Physical education teacher training in fundamental movement skills makes a difference to instruction and assessment practices. *Journal of Teaching in Physical Education*, 34, 548-556.

Lander, N., Morgan, P.J., Solman, J., & Barnett, L.M. (2016). Teachers' perceptions of a fundamental movement skill (FMS) assessment battery in a school setting. *Measurement in Physical Education and Exercise Science,* 20(1), 50-62.

Lander, N., Eather, N., Morgan, P.J., Salmon, J., & Barnett, L.M., (2017). Characteristics of teacher training in school-based physical education interventions to improve fundamental movement skills and/or physical activity: A systematic review. *Sports Medicine*, 47, 135-161.

Leirhaug, P.E., & MacPhail, A. (2015). 'It's the other assessment that is the key': Three Norwegian physical education teachers' engagement (or not) with assessment for learning. *Sport, Education and Society*, 20(5), 624-640.

Lieberman, L., Brian, A., & Grenier, M. (2019). The Lieberman-Brian inclusion rating scale for physical education. *European Physical Education Review*, 25(2), 341-354.

Lima, R.A., Pfeiffer, K.A., Larsen, L.R., Bugge, A., Moller, N.C., Anderson, L.B., & Stodden, D.F. (2017a). Physical activity and motor competence present a positive reciprocal longitudinal relationship across childhood and early adolescence. *Journal of Physical Activity and Heatlh*¸14, 440-447.

Lima, R, A., Pfeiffer, K.A., Bugge, A., Moller, N.C., Anderson, L.B., & Stodden, D.F. (2017b). Motor competence and cardiorespiratory fitness have greater influence on body fatness than physical activity across time. *Scandinavian Journal of Medicine & Science in Sports*,1-10.

Linstone, H.A. & Turoff, M. (2002). The delphi method: Techniques and applications. Massachusetts: Addison-Wesley Publishing Company.

Lopez-Pastor, V.M., Kirk, D., Lorente-Catalan, E, MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18(1), 57-76.

Lorenzo-Seva, U., Timmerman, M.E., & Kiers, H.A. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Res.*, 46(2), 340-364.

Lubans, D.R., Morgan, P.J., Cliff, D., Barnett, L.M., & Okley, A.D. (2010). Fundamental movement skills in children and adolescents: Review of associated benefits. *The Journal of Sports Medicine*, 12(40), 1019–1035

Lund, J. (1992). Assessment and accountability in secondary physical education. *Quest*, 44(3), 352–60.

Lund, J. L., & Veal, M. L. (2008). Measuring pupil learning how do student teachers assess within instructional models? *Journal of Teaching in Physical Education*, 27, 487-511

Lund, L.L., & Kirk, M.F. (2010). Performance-based assessment for middle and high school physical education, second edition. Champaign, Illinois: Human Kinetics.

Lund, J., & Shanklin, J. (2011). The impact of accountability on student performance in a secondary physical education badminton unit. *The Physical Educator*, 68(4), 210-220.

Lund, J., & Tannehill, D. (2005). Standards-based physical education curriculum development. Sudbury, Mass.: Jones and Bartlett Publishers.

MacPhail, A. & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': Students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy & Practice*, 17(1), 23–39.

MacPhail, A., & Murphy, F. (2017). Too much freedom and autonomy in the enactment of assessment. Assessment in physical education in Ireland. *Irish Educational Studies*, 36(2), 237-252.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes. *International Journal of Market Research*, 50, 79–104.

Matanin, M. and Tannehill, D. (1994). Assessment and grading in physical education. *Journal of Teaching in Physical Education,* 13(4), 395–405.

McMillan, S.S., King, M., & Tully, M.P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*, 38, 655-662.

Melograno, V. J. (1997). Integrating assessment into physical education teaching. *The Journal of Physical Education, Recreation & Dance*, 68(7), 34–7.

Michael, R. D., Webster, C., Patterson, D., Laguna, P., & Sherman, C. (2016). Standards-based assessment, grading, and professional development of California middle school physical education teachers. *Journal of Teaching in Physical Education,* 35(3), 277-283.

Mintah, J. K. (2003) Authentic assessment in physical education: prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement, *Measurement in Physical Education and Exercise Science*, 7(3), 161-174.

Morgan, P.J. & Hansen, V. (2007). Recommendations to improve primary school physical education: Classroom teachers' perspective. *The Journal of Educational Research,* 101(2), 99-108.

Morgan, P.J., & Hansen, V. (2008). Classroom teachers' perceptions of the impact of barriers to teaching physical education on the quality of physical education programs. *Research Quarterly for Exercise and Sport*, 79(4), 506-516.

Morrow, J. R., Jr. (1978). Measurement techniques: Who can use them? *Journal of Physical Education & Recreation*, 49, 66-67.

NCATE-National Council for Accreditation of Teacher Education. (2008). Professional standards for the accreditation of teacher preparation institutions. Washington, D.C.: NCATE.

Ni Chronin, D. & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18(2), 219-233.

Oh, J., Graber, K.C., & Woods, A.M. (2016). Are national standards enough?: Convention Supplement. *Research Quarterly for Exercise and Sport*, A109-110.

Pare, G., Cameron, A-F., Poba-Nzaou, P., & Templier, M. (2013). A systematic assessment of rigor in information systems ranking-type Delphi studies. *Information & Management*, 50, 207-217.

Park, Y. (2017). Examining South Korea's elementary physical education performance assessment using assessment literacy perspectives. *International Electronic Journal of Elementary Education*, 10(2), 207-213.

Patton, K. and Griffin, L. L. (2008). Experiences and patterns of change in a physical education teacher development project. *Journal of Teaching in Physical Education*, 27(3), 272–91.

Penney, D. (2012). Points of tension and possibility: Boundaries in and of physical education. *Sport Education and Society*, 18(1), 6–20.

Pickup, I. and Price, L. (2007). Teaching physical education in the primary school – a developmental approach. London: Continuum

Piotrowski, S. (2000). “Assessment, recording and reporting”. *In Teaching physical education* 5–11, Edited by: Bailey, R. and MacFayden, T. 49–61. London: Continuum

Redelius, K. & Hay, P. (2009) Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15(3), 275-294.

Richard, J.F., Godbout, P., Tousignant, M., & Grehaigne, J.F. (1999). The try-out of a team sport performance assessment procedure in elementary and junior high school physical education classes. *Journal of Teaching in Physical Education*, 18, 336-356.

Rink, J. E. (2014). Teaching physical education for learning. Seventh edition. New York, NY.: McGraw-Hill.

Rink, J. E. and Hall, T. J. (2008). Research on effective teaching in elementary physical education. *The Elementary School Journal*, 108(3), 207–18.

Rink, J., Jones, L., Kirby, K., Mitchell, M. and Doutis, P. (2007). Teacher perceptions of a physical education statewide assessment program. *Research Quarterly for Exercise and Sport*, 78(3), 204–15.

Rink, J. & Mitchell, M. (2003). Introduction. State level assessment in physical education: The South Carolina experience. *Journal of Teaching in Physical Education*, 22, 471-472.

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting,* 15, 353-375.

Rowe, G. & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. *Principles of Forecasting,* 125-144.

Safrit, M.J. (1986). Introduction to measurement in physical education and exercise science. Saint Louis, Missouri: Times Mirror/Mosby.

SHAPE America – Society of Health and Physical Educators, (2013). National standards for K-12 physical education. Reston, VA.

SHAPE America. (2013). Grade-Level outcomes for k-12 physical education. Reston, VA.

SHAPE America. (2015). The essential components of physical education. Guidance document. Reston, VA.

SHAPE America. (2016). Every student succeeds act: Game changer for health and physical education. Reston, VA.

SHAPE America. (2017). National standards for initial physical education teacher education (2017). Reston, VA.

Shelton, L.G. (2019). The Bronfenbrenner primer: A guide to develecology. New York, NY: Routledge.

Siedentop, D. and Tannehill, D. 2000. Developing teaching skills in physical education, 4, Mountain View, CA: Mayfield Publishing Company.

Simpson E.J. (1972). The Classification of Educational Objectives in the Psychomotor Domain. Washington, DC: Gryphon House.

Slingerland, M., Borghouts, L., Jans, L., Weeldenburg, G., van Dokkum, G., Vos, S., & Haerens, L. (2017). Development and optimization of an in-service teacher training programme on motivational assessment in physical education. *European Physical Education Review*, 23(1), 91-109.

Sofo, S., Ocansey, R.T., Nabie, M.J., & Asola, E.F. (2013). Assessment practices among secondary physical education teachers in Ghana. *International Online Journal of Educational Sciences,* 5(2), 274-281.

Stiggins, R.J. (1997). Student-centem classroom assessment: 2nd edition. Columbus, OH: Merrill.

Svennberg, L., Meckbach, J., & Redelius, K. (2018). Swedish PE teachers struggle with assessment in a criterion0referenced grading system. *Sport, Education and Society*, 23(4), 381-393.

Thangaratinam, S. & Redman, C.W.E. (2005). The Delphi technique. *The Obstetrician & Gynecologist*, 7 120-125.

Thurlow, M.L. (2002). Positive educational results for all students: The promise of standards-based reform. *Remedial and Special Education,* 23, 195-202.

Tolgfors, B. (2018). Different versions of assessment for learning in the subject of physical education, *Physical Education and Sport Pedagogy,* 23(3), 311-327.

Tousignant, M. & Siedentop, D. (1983). A qualitative analysis of task structures in required secondary physical education classes. *Journal of Teaching in Physical Education,* 47-57.

U.S. Department of Education. (2019). Accreditation in the United States. Retrieved from https://www2.ed.gov/admins/finaid/accred/index.html

van der Mars, H., Timken, G., & McNamee, J. (2018a). Systematic observation of formal assessment of students by teachers (SOFAST). *The Physical Educator*, 75, 431-373.

van der Mars, H., McNamee, J., & Timken, G. (2018b). Physical education meets teacher evaluation: Supporting physical educators in formal assessment of student learning outcomes. *The Physical Educator*, *75(*4), 582–616.

Veal, M.L. (1988). Pupil assessment perceptions and practices of secondary teachers. *Journal of Teaching in Physical Education*, 7, 327-342.

Wardlow, L. (2016). The every student succeeds act in historical context. Retrieved from https://www.pearsoned.com/blog-archive/prek-12-education/

Webster, C.A., Mindrila, D., Moore, C., Stewart, G., Orendorff, K., & Taunton, S. (2019). Measuring and comparing physical education teachers' perceived attributes of CSPAPs: An innovation adoption perspective. *Journal of Teaching in Physical Education*, (ahead of print), 1-14.

Weiss, J. & McGuinn, P. (2016). States as change agents under ESSA. *Kappan,* 97(8), 28-33.

Wiart, L. & Darrah, J. (2001). Review of four tests of gross motor development. *Developmental Medicine & Child Neurology*, 43, 279-285.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14.

Wiliam, D., Lee, C., Harrison, C. and Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49–65.

Wright, M. and Van Der Mars, H. (2004). Blending assessment into instruction. *Journal of Physical Education, Recreation and Dance*, 75(9), 29–35.

APPENDIX A

ELEMENTARY PHYSICAL EDUCATION TEACHER PERCEPTIONS OF MOTOR SKILL

ASSESSMENT SURVEY SUBSCALE ITEMS

Assessment Behavior Items

Item 51: Do you **teach** children **locomotor** skills (i.e., skip, jump, run) in your elementary physical education classes?

a) Yes
b) No

Item 51b: Do you **teach** children **manipulative** skills (i.e., throw, catch, strike) in your elementary physical education classes?

a) Yes
b) No

Item 1: Do you **formally assess** (written/documented) your students' ability to perform **locomotor** skills (i.e., skip, jump, run) in your physical education classes?

a) Yes
b) No

Item 1b: Do you **formally assess** (written/documented) your students' ability to perform **manipulative** skills (i.e., throw, catch, strike) in your physical education classes?

a) Yes
b) No

Motor Skill Assessment Importance Items

Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Somewhat Agree 4	Agree 5	Strongly Agree 6

Item 2: Locomotor skills (e.g., run, jump, leap) should be taught in elementary school physical education.

Item 81: Locomotor skills (e.g., run, jump, leap) should be formally assessed (written/documented) in elementary physical education.

Item 53: Manipulative skills (e.g., throw, catch, kick) should be taught in elementary school physical education.

Item 83: Manipulative skills (e.g., throw, catch, kick) should be formally assessed (written/documented) in elementary physical education.

Motor Skill Assessment Perception Items

To what extent do you agree that the following statements influence your use of formal motor skill assessment?

Strongly Disagree 1	Disagree 2	Somewhat Disagree 3	Somewhat Agree 4	Agree 5	Strongly Agree 6

Item 27: My class sizes are too large to formally assess motor skills as well (quality) as I would like to

Item 27b: My class sizes are too large to formally assess motor skills as frequently (often) as I would like to

Item 67: The time needed to formally assess motor skills negatively impacts the opportunity for my students to learn motor skills in class

Item 30: I typically have enough time to formally assess motor skills

Item 40: I believe that formal motor skill assessment is a good use of my time as a physical education teacher

Item 68: I feel knowledgeable enough about motor skills to formally assess my students

Item 32: I have received adequate training in my teacher certification program to be able to formally assess motor skills

Item 36: I believe that formal motor skill assessment takes place seamlessly in my class, such that it does not disrupt instruction

Item 70: I believe that I have access to the technology that I would need to formally assess motor skills (if I choose to use technology)

Item 41: I believe that it is important to share formal motor skill assessment results with my students so they can use the information to improve their skills

Item 39: My school district administrators (including building principal) support formal assessment of motor skills

Item 72: I believe that physical education should standardize motor skill assessments across the state so that each district in the state administers the same formal motor skill assessments

Item 43: I believe physical education should standardize motor skill assessments across the country so that each state administers the same formal motor skill assessments