EVIDENCE-BASED ASSESSMENT IN ADPATED PHYSICAL EDUCATION-COGNITIVE OUTCOMES: A META-ANALYSIS

By

James Robert Kunkel

A Thesis Presented to

The Faculty of Humboldt State University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Kinesiology: Teaching/ Coaching

Committee Membership

Dr. Rock Braithwaite, Committee Chair

Dr. Chris Hopper, Committee Member

Dr. Sean Healy, Committee Member

Dr. Justus Ortega, Program Graduate Coordinator

July 2017

Abstract

EVIDENCE-BASED ASSESSMENT IN ADPATED PHYSICAL EDUCATION-
COGNITIVE OUTCOMES: A META-ANALYSIS

James Robert Kunkel

There is little data to show evidence-based practices in adapted physical education and whether or not such practices are successful. Currently there is a lack of information on the frequency of which assessments are being administered, on the disabilities that are being assessed or should be assessed during each test, and on the uses for which assessments are being conducted. The aim of this paper is to assess and synthesize all evidence-based practices on cognitive outcomes in adapted physical education using a meta-analysis. Data was sourced from computerized searches using the following databases: SPORT Discus, PsycINFO, PsycARTICLES, Pub Med (Medline), Cochrane Database, Omni File Full Text Mega, ProQuest, Child Development and Adolescent Studies and ERIC. Studies must have been conducted in a physical education/physical activity setting, including children between age 3-22, describe and use assessment practices or intervention in the physical education/physical activity setting, show quantitative statistics and correlations to estimate effect and be conducted between January 1970 and February 2015. The average treatment effect for all evidence-based assessments was small (g = -0.14; *SE* =.13; *95% C.I.* = -0.77, 0.46; $p > 0.05$). Results between subgroups were not significant for any of the subgrouping variables. Overall,

more studies are needed with quantitative data, over longer periods of time, to prove any effectiveness of evidence-based assessments in adapted physical education.

Table of Contents

List of Tables

Introduction

Research in the outcomes of assessment in education have grown over recent years, delivering promising data suggesting that assessment can facilitate learning and improvement through a variety of venues (Sadler, 1998; Black & William, 1998). Similarly, the study of health benefits in fitness and physical education has also been well documented for many decades and has been proven to be positively associated with cognitive performance and academic achievement (Sallis et al., 1998). While research shows that assessment and physical education have their own various individual benefits, they have yet to be studied as one for the use of improving cognition. This paper attempts to synthesize the extant literature, or lack of literature, on the influence of evidence-based assessment on cognitive outcomes in Adapted Physical Education.

Over the past two decades the overall health and fitness levels of individuals with disabilities has become a serious point of emphasis in the education system, inspiring a need to address the individual needs of students with disabilities in Physical Education classes (Murphy & Carbone, 2008; Sallis et al., 1997). Considering the needs of students with disabilities, the federal government enacted legislation to mandate equal opportunity in education leading to the formation of Individuals with Disabilities Education Act (IDEA) (APENS History, 2008). This act required that all students with disabilities have access to physical education in a normal school environment and that each individual receive an Individualized Education Program (IEP) to address the specific and appropriate needs of each individual student. Though students with disabilities are able to participate in general physical activity, often these students qualify for Adapted Physical Education (APE), which is physical education which has been adapted or modified to meet the individual needs of students with disabilities (What is Adapted Physical Education, 2008). APE classes focus on the development of physical and motor skills, fundamental motor skills, and skills in aquatics, dance, and individual and group games and sports (What is Adapted Physical Education, 2008). Individualized educational plans (IEP) use an assessment process to identify goals and track progress towards goals to ensure that student develop the knowledge and skills necessary to live healthy active lifestyles (Torres & Foundation for Exceptional Children, 1977). National standards in physical education have established targets for students and provide a direct for development and though Adapted Physical Education services have been offered for many years empirical evidence is still needed to determine evidence-based practices in assessment that facilitate learning and achievement.

Although there is a lack of evidence-based practices in the specific field of Adapted Physical Education, evidence-based practices are often utilized in other disciplines such as medicine, clinical psychology, school psychology, counseling, behavior analysis, education and nursing. Though there are many varying definitions of the term, Detrich (2008) proposed in a study done at The Wing Institute, that evidence-based practices include three interdependent tasks: identifying, implementing and evaluating interventions with empirical support. Detrich suggested that once an empirically supported intervention is identified, its implementation should be measured to confirm that the intervention is implemented as intended. Once this step is measured, the intervention should then be evaluated to determine whether the intended effects of the treatment were obtained (Reed et al., 2008).

Just as evidence based practices are used in various medical disciplines, methods of evidence-based practices are utilized throughout the education system as well. These approaches are used to create an effective educational practice that is used to teach students of different levels of previous knowledge, allowing all students to achieve, excel and experience their full potential. In order to assist students in reaching their potential, teachers and students need tools and methods that are supported, derived, and understood by research and proven to be successful in the classroom setting. Numerous researchers such the Regional Educational Laboratories (RELs) have set out to provide this information to the public by studying, understanding, developing, applying, and measuring such tools in the educational atmosphere. REL's develop tools and processes through their own research combined with that of others, then collect, analyze, and utilize data to make informed decisions about teaching and learning (North Central Regional Educational Lab at Learning Point Associates, 2003). While some organizations such as this exist to make informed decisions on the broader spectrum of education, it seems that there is a lack of such evidence-based practice research in the field of Adapted Physical Education, specifically in the area of assessment.

Though it has been classified under different terms, assessment has been used in education for many years, and in more recent years has shifted from quantitative assessment techniques for gathering data for improvement in programs or courses to a more qualitative approach. In education, assessment is used generally for gathering and analyzing information about students and using it to improve planning and instruction (Hollandsworth, 1992; Dunn & Mulvenon, 2009). Summative assessment is used as a basis in providing formal, overall assessment at the end of a unit of study and typically inform teachers of a framework for which to base their techniques in instructional content for that particular unit of study (Hollandsworth, 1992; Dunn & Mulvenon, 2009). On the contrary, formative assessments are more student-centered and are used to promote advanced immediate feedback and ultimately produce a more rapid improvement and outcome. In addition, this type of assessment is more informal; allowing teachers to adjust instruction and also to give students more specific and immediate suggestions for improvement while the lesson is in progress (Hollandsworth, L, 1992).

Assessment in Physical Education classes can be used in a multitude of ways including but not limited to motor ability and learning, fitness level improvement, fitness based cognitive learning, and fitness based affective outcomes among others. Physical education assessment specifically has changed rapidly over the years. According to Carroll (1994), around 90% of physical education teachers used Physical Fitness Tests (PFTs) in their programming up until the early 1990's. However, this use of objectives-based assessment has been criticized in more recent research and students often don't understand their meaning and how they apply to real life, resulting in negative experiences. Carroll (1994), did however suggest that such assessment was adequate when used for diagnostic purposes or within self-assessment process, where students can observe their own progress. While PFTs are still an important assessment for individual improvement and for data collection at the state and federal levels, more and more teachers are turning to alternative assessments to assess students' knowledge and ability. In a study reviewing the assessment practices of Physical Education teachers, Desrosiers et al. suggests that alternative assessment, authentic assessment, formative assessment, assessment for learning, and integrated assessment will help move the focus of Physical Education assessment from assessment based on teaching to assessment based on student learning (1997).

Results from recent reviews and meta-analysis indicate that physical activity is positively associated with cognitive performance and academic achievement in children and adolescents (Ardoy et al., 2014). The benefits associated between physical activity cognitive development suggest that improvements occur simultaneously with motor ability due to physical movement that affects the brain's physiology by increasing cerebral capillary growth, blood flow, oxygenation, production of neurotrophins, growth of nerve cells in the hippocampus (center of learning and memory), neurotransmitter levels, development of nerve connections, density of neural network, brain tissue volume, changes in the hormone levels, and increases in arousal (Ardoy et al., 2014). Regular physical activity can improve cognitive function as increases in physical activity might also be associated with improved attention; improved information processing, storage, and retrieval; enhanced coping; and modulation of cognitive control processes to meet task demand (Ardoy et al., 2014). Using this research and the research of Roediger and Karpicke (2006), it is possible that suggest that the use of assessment in APE will help students to retain information about exercise and fitness that will help them continue to exercise as they mature which will in turn help these individuals to sustain and improve their cognitive development.

The purpose of this meta-analysis was therefore to quantify the effect of cognitive outcomes through the use of evidence-based assessment in Adapted Physical Education by including all intervention studies that provided results for assessment in Adapted Physical Education. To our knowledge, this is the first meta-analysis to address this question.

Methods

Search Strategies and Inclusion Criteria

A literature search was conducted by three authors in three separate phases that included a) an electronic database search, b) a search for review articles and c) a search of the reference sections in articles that were included as a part of the screening process. Electronic database searches were performed in *SPORT Discus*, *PsycINFO*, *PsycARTICLES*, *Pub Med (Medline)*, *Cochrane Database*, *Omni File Full Text Mega*, *ProQuest*, *Child Development* and *Adolescent Studies*, and *ERIC* using variations of the keywords *assessment, testing, test, measurement, evaluation, formative assessment, summative assessment, norm-referenced, criterion-referenced*, *affective, cognitive*, *psychomotor*, *mastery learning, rubrics, testing, on-going, and standardized.* Articles retained for the current meta-analysis met the following criteria: (a) Study is conducted in Physical Education/ Physical activity setting in which inclusion of students with disabilities occurs between the age 3-22, (b) describes or uses an assessment practice, method, instrument, or intervention for students during participation in the physical education/ physical activity setting to measure progress, learning, and/or levels of functioning, (c) includes quantitative descriptive statistics and/or correlations to estimate an effect size, and (d) is in the English language and was conducted/published between January 1970 and February 2015.

Coding and Data Extraction

The search strategy included a screening process to select relevant information. Three separate researchers first screened articles by title then by abstract and, when abstracts did not provide sufficient data, the full-text was screened to determine whether inclusion criteria were met. In addition, a screening of reference lists of primary studies, review papers and identified articles were performed as a supplementary search for key terms. To determine whether or not each article met the specific criteria for inclusion, the three separate researchers thoroughly reviewed and coded articles using a coding form.

Coding and data extraction forms following established meta-analytic procedures were used to evaluate and code data to the relevant topic of assessment in Adapted physical education. Information was extracted from each article by three reviewers and included reviewing facts according to three subgrouping categories that included *Methodological Characteristics* 1) Assessment Approach (Formative, Summative, or Both); 2) Assessment Duration (Unit, Semester, Year, or Not Reported); 3) Assessment Setting (Inclusive or Specialized Class); 4) Assessment Focus (Motor, Cognitive, Affective, or Combination), and 5) Assessment Design (Descriptive or Experimental). *Sample Characteristics* included 6) Level of Functioning (Mild, Moderate, or Severe); 7) Environment (Physical Activity, Physical Education, or Sport); 8) Gender (Male, Female, Both); 9) School Level (Elementary, Middle, High or Combination); 10) Study Geographical location (Rural or Urban); 11) Country of Origin (US, UK, etc.); and 12) Parent Support (Parental Support OR No Parent Support). *Study Characteristics* included; 13) Study Measure (Objective or Subjective); and 14) Study Status (Published or Unpublished).

Effect Size Calculations

The Comprehensive Meta-analysis (CMA) Statistical program was employed to compute all effect sizes (Borenstein et. al, 2005). The program provided more than 258 data entry options that were used to calculate effect sizes included variations on both matched and unmatched designs across post-test, pre-post contrast and gain scores. Estimates of effects size calculations were based on descriptive statistics such as means, standard deviations, sample sizes, and when necessary *t* or *p* values (Rosenthal, 1994). When a study reported more than one outcome (multiple outcomes per study), the author chose the study as the unit of analysis which averages outcomes resulting in one overall calculation (Bakeman, 2005). Cohen's *d* was used as the primary measure of effect (Cohen, 1988) and interprets calculations as small ($d \geq 0.20$), moderate ($d \geq 0.50$), or large ($d \geq 0.80$).

Random Effects Model

In a fixed effects model all studies in the meta analysis are thought to share a common effect and differences in effect are a result of sampling error (within study), whereas in a random effects model it is assumed that there is both within study error and between study variance (Hedges & Vevea, 1998). A random effects model was chosen for analyses as there was expected variation between intervention methods, potential sampling error, and the possibility of random unexplained variance between studies (Hedges & Vevea, 1998). Standardized mean differences were adjusted by the inverse weight of the variance to prevent sample size from inflating study weights and allowing for a one accurate calculation of the combined effect size.

Heterogeneity of variance

When employing a random effects model there is a chance that the true effect size will vary between studies, therefore, several indicators were used to assess heterogeneity of variance. The *Q*-statistic is used as a significance test and is based on critical values for chi-square distribution. Significant *Q* values suggest heterogeneity or that the, variability across effect sizes is greater than what would have resulted from chance (Hatala et. al, 2005). Heterogeneous effect size distributions indicate variability that can be explained by study moderators will help provide a more accurate estimate of the distribution.

Outlier Analysis & Publication Bias

An outlier analysis was used to determine if there were any studies that influenced summary effect sizes. If outliers were present a sensitivity analysis ("one study removed" procedure) in CMA was performed by evaluating residual values (*z*-scores). The decision to include potential outliers was based on whether results would remain significant ($p < .05$) and with the 95 percent confidence interval. Publication bias was evaluated using observation of the funnel plot, Trim and Fill procedure (Duval & Tweedie, 2000a; 2000b), and a Fail Safe *N* calculation (Rosenthal, 1979). The funnel plot provides a visual depiction of publication bias with symmetrical plots suggesting lack of publication bias and asymmetrical plots suggest publication bias (Stern, 2001). A Trim and fill procedure adjusts overall effect size by finding the number of studies it would take to provide an unbiased estimate of effect size (Duval, 2006). Fail safe N was used to determine the number of non-significant studies it would take to nullify significant results (Iyengar, 1988)

Results

The primary purpose of the current study was to determine the overall effectiveness of evidence-based practices across all modalities of learning, focusing on cognitive outcomes of assessment and practices for students in the adapted physical education setting. Search procedures generated 300,000 potential studies for evaluation and initial decisions regarding article retrieval were based on review of article titles. After the title screening process, a total of 87,000 were identified as potential sources for data collection. Review of abstracts reduced the potential sources to 16,000 articles. The final screening further decreased total number of articles to 81 for full text review. From the 81 potentially relevant articles six studies met the inclusion criteria, resulting in six independent samples including 249 children or adolescents that were used for analysis. Figure 1 shows the literature search strategy and the primary reasons for exclusion of studies at each stage of the extraction process. Table provides the coding characteristics for each of the article included.

Random Effects Model

The average treatment effect for all evidence-based assessments (across all cognitive outcomes) was small (g = -0.14; *SE* =.13; *95% C.I.* = -0.77, 0.46; $p > 0.05$) and non-significant favoring control groups or conditions. Table 2 presents the overview of the relevant statistics when evaluating the overall effect as there was a significant heterogeneous distribution (QT = 35.11, $p < 0.05$) and that a large portion of variance can be explained ($I^2 = 85.70$) by moderator variables.

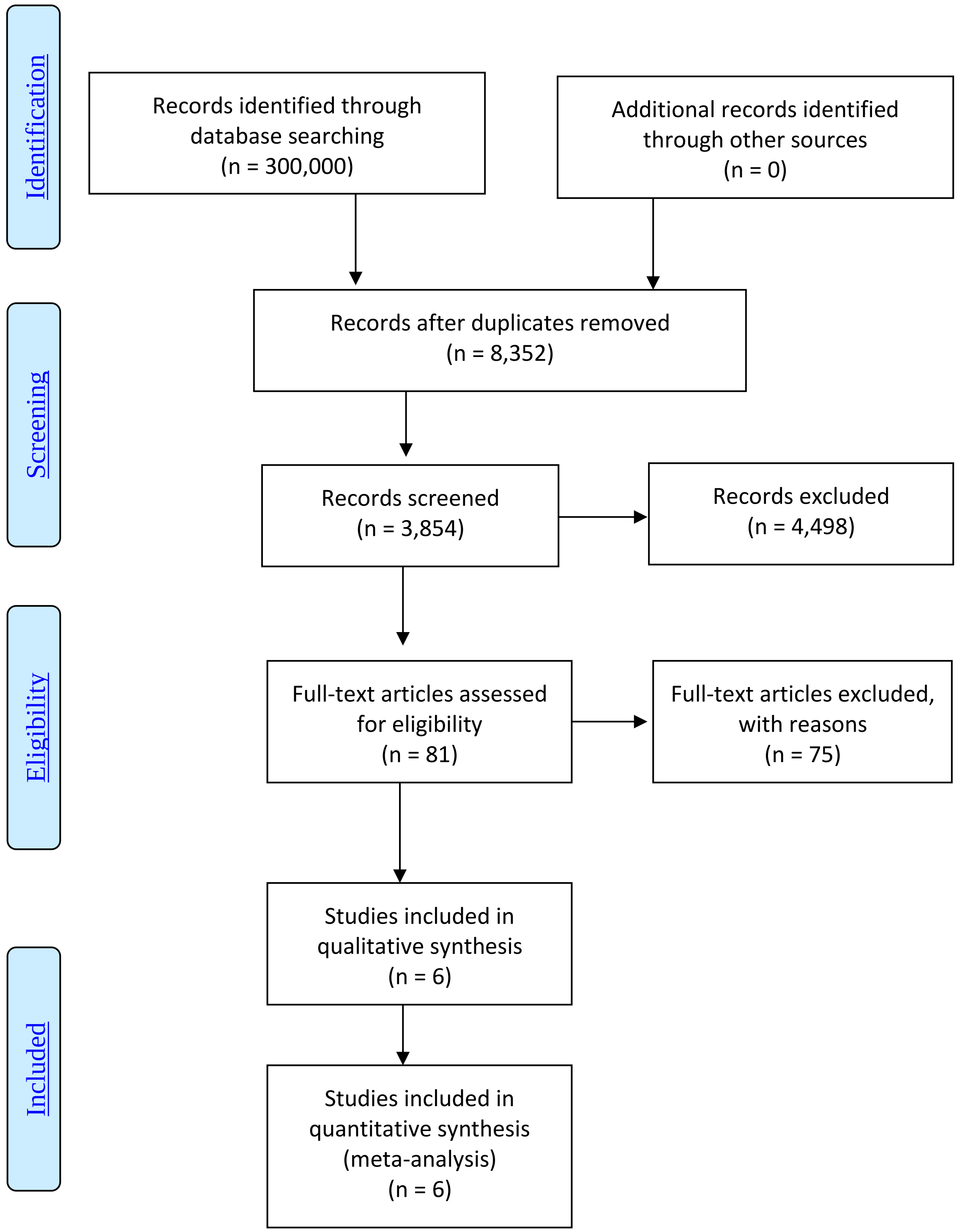

Figure 1. PRISMA Flow Diagram of Literature Search process

Table 1. Coding Characteristics for Studies meeting Inclusion Criteria

	Assessment Characteristics					Participant Characteristics				Study Characteristics	
Study	Approach	Duration	Setting	Focus	Design	N	Level	Gender	Country	Type	Measure
Borremans et al 2009	B	S	S	M	E	20	H	B	Finland	P	C
Peens et al. 2004	S	U	I	M	E	58	E	B	S. Africa	P	C
Shapiro & Dummer 1998	S	U	S	M/C/A	D	50	M	M	US	P	O
Shapiro & Ulrich 2002	F	U	I	C/A	D	60	M	B	US	P	S
Starling 2012	S	S	S	C/A	E	43	M	B	Australia	P	O
Verret et al. 2010	S	S	I	M/C/A	E	18	E	NR	Canada	P	C

Note. Approach = Assessment Approach: F = Formative, S = Summative, B = Both Formative and Summative. Duration = Assessment Duration: U = Unit, S = Semester, and Y = Year. Setting = Assessment Setting: I = Inclusive, S = Specialized Class, O = Other. Focus = Assessment Focus: M = Motor, C = Cognitive, A = Affective, M = Multiple Foci. Design = Assessment Design: D = Descriptive, E = Experimental. Level = Participant Level: E = Elementary, M = Middle School, H = High School, O = Other. Gender = Participant Gender: M = Male Only Class, F = Female Only Class, B = Female and Male Class. Type = Study Type: P = Published, U = Unpublished. Measure = Study Measures: S = Self-Report, O = Objective, C = Combined Self-Report and Objective.

Table 2. Subgroup Analyses

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
	k	*g*	*SE*	s^2	95% C.I.	*Z*	*Q*	τ^2	I^2	Fail Safe N
Random Effects Model [a]	6	-0.14	0.32	0.10	(-0.765, 0.483)	-0.44	35.11	0.48	85.76	330
Methodological Characteristics [b]										
Assessment Approach							5.06[b]			
Both	1	.26	.75	.51	(-1.139, 1.666)	0.3.68	0.00	0.00	0.00	
Formative	2	.54	.44	.19	(-0.310, 1.399)	1.25	2.03	0.07	50.61	
Summative	3	-0.72	.38	.15	(-1.475, 0.031)	-1.88	11.83*	0.48	83.09	
Assessment Duration							1.10[b]			
Unit	3	-0.57	0.53	0.28	(-1.614, 0.466)	-1.08	21.36*	0.49	90.64	
Semester	3	0.17	0.47	0.23	(-0.758, 1.099)	0.36	13.46*	1.17	85.14	
Assessment Setting							1.10[b]			
Inclusive	3	-0.1	0.12	0.01	(-0.329, 0.137)	-0.81	21.36	0.49	90.64	
Specialized	3	-	0.24	0.01	(-0.702,	-1.01	13.46	1.17	85.14	

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
	k	*g*	*SE*	s^2	95% C.I.	*Z*	*Q*	τ^2	I^2	Fail Safe N
		0.24			0.225)					
Assessment Focus							0.27[b]			
Motor	1	0.26	0.86	0.74	(-1.424, 1.951)	0.31	0.00	0.00	0.00	
Multiple	5	-0.22	0.36	0.13	(-0.931, 0.487)	-0.61	34.32*	88.35	88.35	
Assessment Design							0.34[b]			
Descriptive	1	0.28	0.83	0.69	(-1.348, 1.908)	0.34	0.00	0.00	0.00	
Experimental	5	-0.26	0.39	0.15	(-1.025, 0.509)	-0.66	32.19*	0.62	87.57	
Sample Characteristics[b]										
Age							5.78[b]			
Elementary	2	-1.22	0.56	0.31	(-2.320, -0.122)	-2.18*	8.42*	1.78	88.12	
Middle	1	0.00	0.74	0.55	(-1.453, 1.453)	0.00	0.00	0.00	0.00	
High	1	0.26	0.81	0.66	(-1.326, 1.853)	0.33	0.00	0.00	0.00	
Multiple	2	0.55	0.51	0.26	(-0.460, 1.550)	1.06	35.11	0.07	50.61	
Country							19.46[b]			
Australia	1	0.00	0.41	0.17	(-0795,	0.00	0.00	0.00	0.00	

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
	k	*g*	*SE*	s^2	95% C.I.	*Z*	*Q*	τ^2	I^2	Fail Safe N
					0.795)					
Canada	1	-2.54	0.73	0.53	(-3.957, -1.113)	-3.50*	0.00	0.00	0.00	
Finland	1	0.26	0.52	0.27	(-0.761, 1.288)	0.50	0.00	0.00	0.00	
S. Africa	1	-0.53	0.31	0.10	(-1.133, 0.077)	-1.71	0.00	0.00	0.00	
US	2	0.54	0.27	0.07	(0.021, 1.062)	2.04*	2.03	0.07	0.00	
Study Characteristics [b]							38 [b]			
Measure										
Self-Report	1	0.28	0.99	0.98	(1.657, 2.217)	1.08	0.00	0.00	0.00	
Objective	1	0.00	1.00	1.03	(-1.962, 1.962)	0.00	0.00	0.000	0.00	
Combination	4	-0.37	0.52	0.27	(-1.381, 0.650)	-1.94	31.66*	0.909	90.62	

Note. *k* = number of effect sizes. *g* = effect size (Hedges g). *SE* = standard error. S2 = variance. 95% *C. I.* = confidence intervals (lower limit, upper limit). *Z* = test of null hypothesis. τ^2 = between study variance in random effects model. I^2= total variance explained by moderator. * indicates p < .05. a = Total *Q*-value used to determine heterogeneity. b = Between *Q*-value used to determine significance (α < 0.05).

Moderator Analyses

There was a significant heterogeneous distribution and moderator (Subgroup) analyses, however, given that the confidence interval was both positive and negative results were not tenable. Summary information for each moderator category is reported below.

Methodological Characteristics. Of the 6 studies reviewed, three were summative assessments, two were formative, and one was a combination of both formative and summative assessments. Of these six, assessment durations were split, with three unit long and three semester long assessments. Assessment settings were also split at three between inclusive and specialized settings. Of the reviewed studies, only one focused solely on motor ability while the remaining five focused on cognitive ability with a combination of either motor and/or affective ability. Assessment designs were mostly experimental (5 assessments), while one was a descriptive design.

Sample Characteristics*.* Studies included between 18 and 60 participants, and were conducted with both male and female participants. Four studies included both boys and girls, while one used only males, and another did not report the gender of its participants. Participants were between elementary school age and high school age with 2 samples being from elementary school, one from middle school, one from high school, and two samples being a combination of the age groups. The samples were diverse in origin as they came from a variety of countries including two samples from the United States, one from Australia, one from Canada, one from Finland, and one from South Africa.

Study Characteristics. Of the 6 included studies, all articles were published and only two experiments used solely a self-report form of measurement or an Objective form of measurement. The remaining four studies used a combination both self-reported measures as well as objective measures.

Outcome Analysis

Outcome analyses were not conducted as no outcome was reported more than once, preventing any interpretation of results. The discussion section provides plausible explanations for the lack of findings.

Discussion

This meta-analysis was conducted to give an overview and a culmination of studies and data that have performed evidence-based cognitive assessments in the field of Adapted Physical Education. Results indicated that there were no overall positive effects and no significant outcomes. Studies in our research reported on outcomes including Perceived Confidence, Language, Listening Comprehension, Cognitive Functioning, (i.e., processing speed), Perceived Importance, Usefulness, and Enjoyment; however, no study reported on more than one outcome and no significant data was provided. Therefore, no knowledge was reported. Although none of the effect sizes were large enough to be significant, our study conducted a moderator analysis to help indicate some of the possible factors that may have influenced the effectiveness of each intervention.

Assessment Approach

The majority of the studies included within this meta-analysis were based on summative assessment, in which information was collected at the end of a unit to evaluate student learning as compared against a standard or benchmark. Our Results show that these assessments showed a moderate to large negative effect size, indicating that students with disabilities performed at a lower level on summative assessments than did their typically developing peers. This may be explained through data collection, as in the process of summative assessment, data is only collected at the conclusion of the study, leaving no room for progressive decision-making, student improvement, or changes throughout the study. While such assessment strategies are useful in determining the overall success of an intervention, it should be expected that in most cases students without disabilities would progress and learn more over a longer period of time than will students without disabilities. In the case of formative assessment, however, results showed a moderate positive effect size, suggesting that this type of assessment was more useful for students with disabilities than it was for their typically developing peers. This style of assessment is important in monitoring the progression of individuals within an intervention period and provides opportunity for feedback and improvement (Andrade & Cizek, 2010). Formative assessment is important to implement, helping monitor individual learning and providing ongoing feedback while also helping to make decisions and guide future learning based off current performance. This positive effect size goes hand in hand with current research that shows formative assessment helps to improve the learning of low-achieving students, including those with disabilities, even more than it helped other students (Black and William, 1998). The evidence that is documented from this type of assessment is important not only to the success of the individual but also to help guide decisions in Adapted physical education.

Assessment Duration

Each of the studies included in our analysis conducted interventions in short units or semester long durations of study. While some interventions can show progression in a short time period, others may take longer to show such progression. As the moderate negative effect size shows, students with disabilities performed at a lower level than their peers without disabilities, and performed better on lengthier assessment periods. With the use of lengthier studies, interventions can prove to be successful where individuals in shorter studies were unable to gain the experience necessary to reach the same success. As explained in previous studies, students with intellectual disabilities learn at a much slower rate (Vaughn, Bos, Schumm, 2007) and take longer to process and obtain information (Bennet, 1997). This could be a valid explanation for the duration effect sizes we found in our current study.

Assessment Focus

The majority of assessments used in the included studies were based on a combination of psychomotor skills, cognitive skills, and affective skills. As the effect size shows, students with disabilities were out performed by their peers without disabilities. Effect sizes in assessment focus also determined that students with disabilities showed more success during the motor based assessments. Such a small effect size may be explained by the participant's level of functioning. Because most students with disabilities are already at a low level of function, any improvement or declines no matter how small, can seem dramatic, as they can be both immediate and noticeable. This data is exemplified in a study conducted by Rose et al., demonstrating that oxygen uptake and heart rate elevated higher and at a slower walking speed for children with cerebral palsy than it did for normal children (1990). While the data retrieved from these combinations of skills is a great achievement and could be used extensively to make decisions regarding assessment in adapted physical education, data focused singularly on cognitive outcomes could be a beneficial addition to the direction and success of individuals in adapted physical education.

Assessment Design

Of the six included studies in this meta-analysis, all but one used experimental designs in which a casual connection between the independent and dependent variables is established (Millsap & Maydeu-Olivares, 2009). In the case of the five articles using experimental designs, regularly developing students were compared against students with disabilities. As shown by the effect size, the study showed that students with disabilities were out performed by their regularly developing peers. While this data displays a gap in skills between the two groups, it fails to determine what assessments are more successful for students with disabilities.

Sample Age

In regards to the sample age of our meta-analysis, elementary aged students produced a small negative effect size, indicating that students with disabilities in elementary school performed at a lower level than their regularly developing peers. This effect size may be a result of experience in assessment. As explained by The Early Childhood Assessments Resource Group, young children do not have the experience to understand what the goals of formal testing are, therefore testing interactions may be very difficult or impossible to structure appropriately (Shepard, et al., 1998). For example, the majority of the studies used in our meta-analysis incorporated self-report methods, which may be unreliable for children at younger ages as it is difficult for young children to recall physical activity behavior due to sporadic activity patterns and short duration of sessions (Mattocks et al. 2008). Additionally, children often overestimate the intensity and amount of time being physically active (Hussey et al. 2007). Considering how difficult these tasks can be for regularly developing children, it’s fair to say that students with disabilities would struggle even more on the same tasks, which would explain the negative effect size.

The sample age also revealed that multiple age levels had a moderate positive affect, indicating that students with disabilities that were diverse in age performed better than their typically developing diverse aged peers. This effect size may be explained by the increased opportunities to work with peers of multiple ages and benefit from a range of learning strategies, including modeling and peer mentoring. Such strategies could dramatically effect the outcome and increase the success levels of either group, but may have a larger effect on those with disabilities.

Sample Country

The studies used in this meta-analysis were pulled from a diverse group of countries including Austrailia, Canada, Finland, South Africa, and the United States. Using this diverse group produced a wide variety of effect sizes which could be explained by a multitude of factors including cultural differences and laws and regulations. Being such a diverse group of countries, it is expected that each country has its own cultural identity which could have many different positive and negative affects on the learning focus as well as the curriculum that is used in each study. These cultural differences could help explain the differences in effect size between each country. In regards to laws and regulations it should also be expected that each country has its own set of laws to govern education. These laws can affect the amount and types of resources that are provided. Resources such as teacher training, time allotted for learning, student support, and various other factors could have a large impact on the success of student learning during assessment.

Overall, the meta-analysis shows that there is little quantitative data to show that evidence-based practices in adapted physical education are successful, specifically those used for cognitive outcomes. Results indicated that there were no overall positive effects and no significant outcomes, this is likely due to the fact that no study reported on more than one outcome, providing no significant data to our research. Future research should focus on filling the gaps identified in this review, such as assessment durations and approaches while targeting both adapted physical education and cognitive outcomes. Future studies should aim to strengthen the evidence base for interventions among adolescent boys and girls with rigorous designs, longer follow-ups, use of objective measures, and assessment of cognitive outcomes.

References

APENS History. (2008). Retrieved July 10, 2017, from http://apens.org/history.html

Ardoy, D. N., Fernández-Rodríguez, J. M., Jiménez-Pavón, D., Castillo, R., Ruiz, J. R. and Ortega, F. B. (2014), A Physical Education trial improves adolescents' cognitive performance and academic achievement: the EDUFIT study. Scand J Med Sci Sports, 24: e52–e61. doi:10.1111/sms.12093

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, *37*(3), 379-384.

Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1): 7–74.[Taylor & Francis Online],

Borenstein, M, Hedges, L, Higgins, J, Rothstein, H. (2005). Comprehensive meta-analysis version 2. *Englewood, NJ: Biostat*, 104.

Borremans, E., Rintala, P., & McCubbin, J. A. (2009). Motor skills of young adults with Asperger syndrome: A comparative study. *European Journal of Adapted Physical Activity*, 2 (1), 21-33. Retrieved from http://eujapa.upol.cz/index.php/EUJAPA/article/view/12/11

Carroll, B. (1994) Assessment in Physical Education. Ataccher´s guide to the Issues (London, Falmer Press).

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Desrosiers, P., Genet-Volet, Y. & Godbout, P. (1997) Teachers’ Assessment Practices Viewed Through the Instruments Used in Physical Education Classes, Journal of Teaching in Physical Education, 16(2), 211-228.

Detrich, R. (2008). Evidence-based, empirically supported, or best practice? A guide for the scientist-practitioner. In J. K. Luiselli, D. C. Russo, W. P. Christian, & S.Wilczynksi (Eds.): *Effective practices for children with autism: Educational and behavior support interventions that work*. New York: Oxford Press.

Dunn, Karee E and Mulvenon, Sean W. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. Practical Assessment Research & Evaluation, 14(7). Available online: http://pareonline.net/getvn.asp?v=14&n=7

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics*, *61*(1), 103-131.

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89-98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.

Hatala, R., Keitz, S., Wyer, P., Guyatt, G., & Evidence-Based Medicine Teaching Tips Working Group. (2005). Tips for learners of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. *Canadian Medical Association Journal*, *172*(5), 661-665.

Hedegs, L.V., & Vevea, J.L. (1998). Fixed and random effects models in meta-analysis. New York: Russell Sage Foundation.

Hollandsworth, L. (1992). Understanding assessment: Guidelines for utilizing a qualitative approach. Eric Document Reproduction Services, No.ED 353 386.

Iyengar, V., & Woittiez, J. (1988). Trace elements in human clinical specimens: evaluation of literature data to identify reference values. *Clinical chemistry*, *34*(3), 474-481.

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. Doi:10.1080/13573322.2012.713860

Millsap, R.E., & Maydeu-Olivares, A. (2009). SAGE Handbook of Quantitative Methods in Psychology. Sage Publications Ltd. doi: http//dx.doi.org/10.4135/9780857020994

Murphy, N. A., & Carbone, P. S. (2008). Promoting the Participation of Children With Disabilities in Sports, Recreation, and Physical Activities. Retrieved July 10, 2017, from http://pediatrics.aappublications.org/content/121/5/1057.short

North Central Regional Educational Lab at Learning Point Associates (2003). *Evidence-based Education: Bridging Research and Practice for Student Success*, (6). http://files.eric.ed.gov/fulltext/ED483176.pdf

Peens, A. A., Pienaar, A. A. E., & Nienaber, A. A. W. (2004). THE EFFECT OF DIFFERENT INTERVENTION PROGRAMMES ON THE SELF-CONCEPT AND MOTOR ABILITY OF 7-9 YEAR OLD DCD CHILDREN. (Abstract) *In Klisouras, V. (ed.), 2004 PRE-OLYMPIC CONGRESS: Proceedings: VOLUME I: LECTURES-ORALS: SPORT SCIENCE THROUGH THE AGES, 6-11 AUGUST*

2004, THESSALONIKI/HELLAS, ARISTOTLE UNIVERSITY CAMPUS, Thessaloniki, Greece, Aristotle University of Thessaloniki, Department of Physical Education & Sport Science, [2004], p.445-446.;.

Reed, F. D. D., & Reed, D. D. (2008). Towards an understanding of evidence-based practice. *Journal of Early and Intensive Behavior Intervention, 5*(2), 20-29. http://dx.doi.org/10.1037/h0100416

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: basic Research and Implications for Educational Practice. *Perspectives on Psychological Science, 1*(3), 181-210. doi:10.1111/j.1745-6916.2006.00012.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*(3), 638.

Rosenthal, R. (1994). Statistically describing and combining studies. In H. Cooper & L. Hedeges (Eds.), The handbook of research synthesis (pp. 231-244). New York: Russell Sage Foundation.

Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education*, 5(1): 77–84.[Taylor & Francis Online],

Sallis, J.F., McKenzie, T.L., Alcaraz, J.E., Kolody, B., Faucette, N., & Hovell, M.F. (1997). Effects of a two-year health-related physical education program on physical activity and fitness in elementary school student: SPARK. American Journal of Public Health, 87, 1328-1334.

Sallis, J. F. Mckenzie, T. L., Kolody, B., Lewis, M., Marshall, S., & Rosengard, P. (1999). Effects of Health-Related Physical Education on Academic Achievement:

Project SPARK. *Research Quarterly for Exercise and Sport, 70*(20), 127-134. Doi:10.1080/02701367.1999.10608030

Shapiro, D.R. and G.M. Dummer, *Perceived and actual basketball competence of adolescent males with mild mental retardation.* Adapted Physical Activity Quarterly, 1998. **15**(2): p. 179-190.

Shapiro, D.R. and D.A. Ulrich, *Expectancies, values, and perceptions of physical competence of children with and without learning disabilities.* Adapted Physical Activity Quarterly, 2002. **19**(3): p. 318-333.

Starling, J., et al., *Training Secondary School Teachers in Instructional Language Modification Techniques to Support Adolescents with Language Impairment: A Randomized Controlled Trial.* Language, Speech, and Hearing Services in Schools, 2012. **43**(4): p. 474-495.

Torres, Scottie & Foundation for Expectional Children (1977). *A Primer on individualized education programs for handicapped children.* The Foundation, Reston, Va

What is Adapted Physical Education? (2008). Retrieved July 10, 2017, from http://apens.org/whatisape.html

Verret, C., P. Gardiner, and L. Béliveau, *Fitness Level and Gross Motor Performance of Children With Attention-Deficit Hyperactivity Disorder.* Adapted Physical Activity Quarterly, 2010. **27**(4): p. 337-351.