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

By

Breanna Lee Ketel

A Thesis Presented to

The Faculty of Humboldt State University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Kinesiology: Exercise Science

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-AFFECTIVE OUTCOMES: A META-ANALYSIS

Breanna Lee Ketel

Currently there is a lack of evidence about assessment in adapted physical education (APE) settings concerning the justification for methods or curricula being implemented by teachers in their classrooms This is perhaps due to a lack of understanding of the central notion of evidence-based practices (Jin & Yun, 2010). The purpose of the current investigation was to conduct a meta-analytic review that evaluates assessment practices to determine the overall effect on specific student affective outcomes. A secondary purpose was to evaluate the moderating effects of different methodological, sample, and study variables. 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. The average treatment effect for all evidence-based assessments (across all affective outcomes) was small (g = -0.43; SE =.24; *95% C.I.*= -0.89, 0.04; $p > 0.05$) and non-significant favoring control groups or conditions. There was a significant heterogeneous distribution for affective outcomes and moderator (Subgroup) analyses, however, given that the confidence interval was both positive and negative results are not tenable. As a result of the findings, more research, with quantitative data, needs to be done to prove the effectiveness of evidence-based assessments in adapted physical education.

Table of Contents

List of Tables

List of Figures

Introduction

Adapted Physical Education (APE)

Federal law states that every child has the right to a "free and appropriate education." For students with disabilities, the word "appropriate" becomes a bit more complex, especially in a physical education context. Adapted Physical Education, as defined by the Adapted Physical Education National Standards (Kelly, 2006), is "physical education which has been adapted or modified so that it is as appropriate for the person with a disability as it is for a person without a disability." This means providing the student with all of the necessary tools he or she needs to be successful in a physical education setting, as well as to develop the skills needed to remain physically active/fit throughout their lives. The Individuals with Disabilities Education Act IDEA (1990) categorizes the disabilities that qualify students to receive special education services. These categories are, 1) autism; 2) deaf-blindness; 3) deafness; 4) hearing impairment; 5) intellectual disability; 6) multiple disabilities; 7) orthopedic impairment; 8) other health impairment; 9) serious emotional disturbance; 10) specific learning disability; 11) speech or language impairment; 12) traumatic brain injury; [and] 13) visual impairment including blindness" (Kelly, 2006). Because it is a federally mandated law that students with disabilities be provided physical education, the adapted physical education teacher is considered a "direct service provider" and must provide physical education as part of the students' special education services (https://www.apens.org/whatisape.html). Federal law defines physical education as, 1) the development of physical and motor skills; 2) fundamental motor skills and patterns

(throwing, catching, walking, running, etc.); 3) skills in aquatics; 4) dance; and 5) individual and group games and sports (including intramural and lifetime sports)" (Kelly, 2006). To accomplish this, and to best meet the needs of the student with a disability, adapted physical educators must use evidence-based practices to guide curriculum development and instruction, to assess student learning, and to achieve desired affective outcomes in their classrooms (Jin & Yun, 2013).

Evidence-Based Practice

Research shows that current practice in adapted physical education is derived from areas such as intuition rather than backed by scientific research and evidence-based practice. According to Philip Davies (1999), from the *Department for Continuing Education* at *University of Oxford*, evidence based practice operates at two levels. The first is to, "utilize existing evidence from worldwide research and literature on education and associated subjects," and the second is to, "establish sound evidence where existing evidence is lacking or of a questionable, uncertain, or weak nature" (Davies, 1999). In the field of adapted physical education, this becomes an increasingly important practice as students are becoming less active and the health consequences present challenges as well as the fact assessments fail to prove that student learning has actually occurred.

Assessment

Summative Assessment. Traditionally, assessment has been used to quantify and analyze student knowledge on a topic at the end of a lesson/unit, and to guide further instruction based on its results. Evidence shows us, however, that this method may not be the most effective for actually gauging student learning (Harlen & Deakin, 2002). By the

end of a lesson/unit in adapted physical education, it may be too late for a student to effectively make adjustments to a newly learned motor skill because they have been practicing it wrong throughout the lesson/unit. Often times, summative assessment also fails to tell us that student learning as even occurred (Lahey, 2014). Though the student may "pass" the test, it is unknown as to whether or not the student will retain such knowledge and if it will help them in maintaining lifetime fitness when they are no longer part of a structured physical education program (Lahey, 2014).

Formative Assessment. Research suggests that students learn best when assessment occurs on a continuous basis and is followed by instant, constructive feedback (Jones, 2005). Haug (2015) states that, "formative assessment, and especially feedback, is considered essential to student learning [and] to provide effective feedback teachers must act upon the information that students reveal during instruction. Using formative assessment in APE means consistently providing students with feedback on ways in which they can improve, while also providing them with praise for things in which they are doing correctly (Haug, 2015). In this use of positive reinforcement and constructive criticism, students can make adjustments before a skill/lesson is leaned the incorrect way. For example, when instant feedback is provided to a student who is failing to step with the opposing foot as he/she approaches to kick a soccer ball, the student can make adjustments before the student develops muscle memory of an incorrect motor skill. Formative assessment has been shown to improve students' cognitive, psychomotor, and affective outcomes (Nicol & Macfarlane-Dick, 2007). Though it is proven that this method of assessing students yields positive results and is linked directly to student

learning, research is lacking that it is actually being practiced, effectively, in adapted physical education settings. Not only does consistent, positive feedback lead to student learning, it also models to students how to effectively communicate with one another to create a classroom environment that strives towards positive affective outcomes (Haug, 2015).

Affective Outcomes

Teachers of physical education have a responsibility to focus on the following areas: 1) health-related physical fitness; 2) cognitive domain; 3) psychomotor domain; and 4) affective domain (Hansen, 2008). In a study on affective outcomes in physical education, Bertelsen (2002) states that, "the implementation of appropriate teaching practices in physical education can contribute to increasing the quality and value of physical education outcomes." Often over looked by educators, affective outcomes of students in adapted physical education, becomes increasingly important to focus on as students with disabilities are integrated with their able bodied peers, in an inclusive classroom setting (Bertelsen, 2002). Affective outcomes should be developed and fostered in school so that the student will learn how to establish and maintain positive interactions and relationships throughout their lives. These positive interactions and relationships are the foundation to the ultimately bigger picture of them maintaining lifetime fitness and health (Bertelsen, 2002). The National Association for Sport and Physical Education (NASPE) defines national standards for physical education. Of these standards, NASPE states that a *physically educated person* "exhibits responsible personal and social behavior that respects self and others in physical activity settings" and "values

physical activity for health, enjoyment, challenge, self-expression, and/or social interaction" (NASPE, 2011). This is where evidence-based practice and appropriate assessment become an imperative part of the adapted physical education curriculum. Though we do not need research to tell us that students with disabilities are bullied and encounter negative peer interactions more so than students without disabilities, it is evidence-based research that informs professionals on the most effective ways in which to teach our students empathy, positive communication skills, and to foster these relationships amongst our students.

Hypothesis

Assessment is a process by which teachers use evidence about student learning and performance in their decision making to facilitate meaningful change. Currently there is a lack of evidence about assessment in APE settings concerning the justification for methods or curricula being implemented by teachers in their classrooms, that is perhaps due to a lack of understanding of the central notion of evidence-based practices (Jin & Yun, 2010). APE teachers need to use assessment to determine the needs of students with disabilities as there is a 40% prevalence of overweight and obesity (Einarsson et al., 2015) . Given these facts there is an imperative for students with disabilities to have opportunity and access to structured daily physical education that uses- evidence to support the decisions that are being made about the activities and instruction being implemented. The use of evidence-based assessments in adapted physical education are lacking information of the frequency that assessments are used, on the disabilities that are being assessed, and the uses of the data obtained from assessments. The purpose of the

current investigation was to conduct meta-analytic review that evaluates assessment practices to determine the overall effect of specific student affective outcomes. A secondary purpose what to evaluate the moderating effects of difference methodological, sample, and study variables.

Methods

Search Strategy & Inclusion Criteria

A literature search was conducted 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.* Three authors conducted the search process in three separate phases that included review of titles to sort literature findings, followed by review of title and abstracts, and then full text retrieval to make final decisions. Figure 1 provides the screen form used to make final decisions after the full text retrieval was completed. 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.

Author & Year:_____ Today's Date:_____

Study Number:_____ Reveiwer:_____

Question	Yes	Not Clear	No	Further information:
Involved in PE/PA /Sport setting?				Which Setting?
Were the participants ages 3 to 22 years?				Average Age?
Did the study implement an assessment, method or intervention in PE/PA/Sport?				Describe Method or Intervention?
Has at least one outcome (quantitative measure) been assessed and reported on**?**				If NO, is an outcome measure related to learning in any way? State the primary measure reported:
English language?				
Published after the year 1970?				
Population is identified with a disability or special need?				
IF THE ANSWER TO ANY OF THE ABOVE IS SHADED BOX, EXCLUDE THE STUDY (FROM THIS INITIAL SCREENING)				
This study is:	**Included**	☐	**Excluded**	☐ **Not sure** ☐
	Details:			
Other information				

Figure 1. Assessment in Adapted Physical Education Meta-Analysis Screening Form

Coding & Data Extraction

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). Figure 1 provides the screening form

Effect Size Calculations

The Comprehensive Meta-analysis (CMA) Statistical program was employed to compute all effect sizes (BioStat, 2014). 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 (Valentine et. al, 2003). 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, 2005). Heterogeneous effect size distributions indicate variability that can be explained by study moderators will help provide a more accurate estimate of the distribution.

Publication Bias & Outliers

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 & Tweed, 2000; 2001), and a Fail Safe *N* calculation (Rosenthal, 1981). 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 (Ivengar, 1988).

Results

The main purpose of this of this meta-analysis project was to compile a collection of data and research that supports effective ways in which to assess students with disabilities, with a focus on the affective outcome domain of physical education, based on evidence in the field of adapted physical education. The search produced 8352 written works with titles that were potentially relevant to the study. Of these 8352 titles, a total of 42 studies met the inclusion criteria, and 4 specifically contained affective outcomes. Figure 1 displays the literature search results and Table 1 shows the coding characteristics for studies that were included in the current analysis.

Random Effects Model

The average treatment effect for all evidence-based assessments (across all affective outcomes) was small (g = -0.43; SE =.24; *95% C.I.*= -0.89, 0.04; $p > 0.05$) 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 ($Q_T = 6.85$, $p > 0.05$) and that a large portion of variance can be explained ($I^2 = 56.21$) by moderator variables.

Subgroup Analysis

There was n significant heterogeneous distribution for affective outcomes and moderator (Subgroup) analyses, however, given that the confidence interval was both positive and negative results and not tenable. Summary information for each moderator category is reported below. Table 2 provides the moderator statistics for all studies reporting affective outcomes.

Methodological Characteristics. Four of 40 studies were included in the affective outcomes subgroup of this meta-analysis. Of these 4 studies, 1 was formative while 3 were summative. Two of the studies took place over an entire semester, and 2 the duration of a unit. Two of the studies took place in an inclusive setting and 2 were conducted in specialized settings. One of the studies followed a descriptive assessment design, and 3 were experimental.

Sample Characteristics. Two of the studies took place in an elementary school setting (elementary aged children), 1 in a middle school setting (middle school aged children), and 1 occurred in a combination of the two settings. The 4 studies took place in different countries including Australia, Canada, South Africa and the United States.

Study Characteristics. One of the studies used self-reported measurements. Another study used objective measures while the two remaining studies used a combination of the two.

Outcome Analysis

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

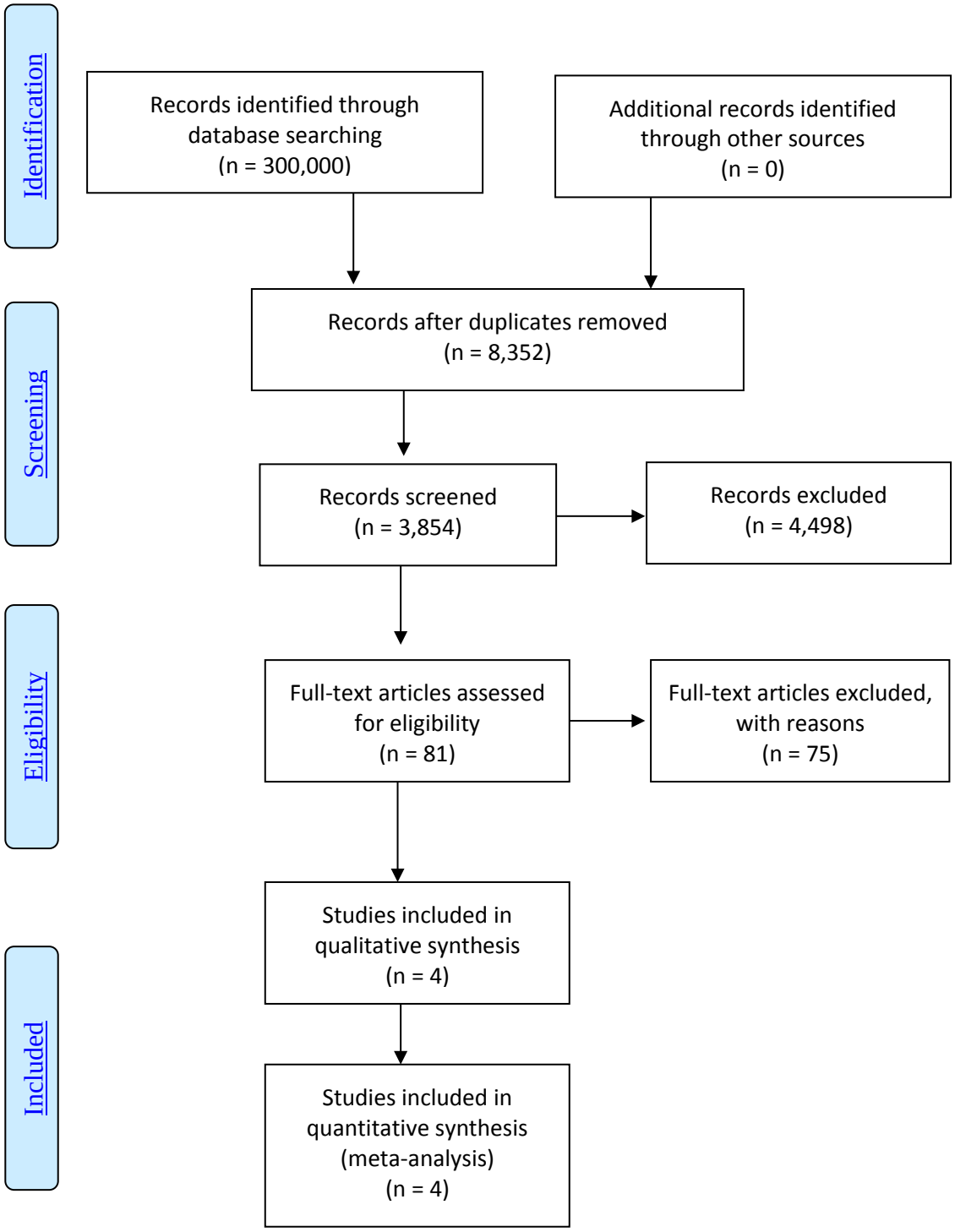

Figure 2 PRISMA Flow Diagram of Literature Search process

Table 1. Coding Characteristics for Studies meeting Inclusion Criteria

	Intervention Characteristics				Sample Characteristics					Study Characteristics	
Study	Approach	Duration	Setting	Focus	Design	N	Level	Gender	Country	Type	Measure
Peens et al. 2004	S	U	I	MULT	E	58	E	B	S. Africa	P	C
Shapiro & Dummer 1998	S	U	S		D	50	M	M	US	P	O
Slaman et al. 2014[	S	S	S	M	E	37	H	NR	Netherlands	P	O
Verret et al. 2010	S	S		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. Moderator Analysis

	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]	4	-0.425	0.235	0.055	(-0.885, 0.035)	-1.813	6.850	0.118	56.2	330
Methodological Characteristics [b]										
Assessment Approach							0.030			
Formative	1	-0.206	0.555	0.308	(-1.294, 0.883)	-0.371	0	0	0	
Summative	3	-0.566	0.350	0.123	(-1.252, 0.121)	-1.616	6.303	0.241	68.2	
Assessment Duration							0.241			
Unit	2	-0.340	0.380	0.144	(-1.084, 0.405)	-0.894	0.561	0	0	
Semester	2	-0.628	0.449	0.202	(-1.508, 0.252)	-1.399	6.265*	1.077	84.0	
Assessment Setting							0.241			
Inclusive	2	-0.340	0.380	0.144	(-1.084, 0.405)	-0.894	0.561*	0	0	
Specialized	2	-0.628	0.449	0.202	(-1.508, 0.252)	-1.399	6.265*	1.077	84.0	
Assessment Design							0.301			
Descriptive	1	-0.206	0.555	0.308	(-1.294, 0.883)	-0.371	0	0	0	
Experimental	3	-0.566	0.350	0.123	(-1.252, 0.121)	-1.616	6.303*	0.241	68.2	
Assessment Approach							0.030			
Formative	1	-0.206	0.555	0.308	(-1.294, 0.883)	-0.371	0	0	0	
Summative	3	-0.566	0.350	0.123	(-1.252, 0.121)	-1.616	6.303	0.241	68.2	
Assessment Duration							0.241			
Unit	2	-0.340	0.380	0.144	(-1.084, 0.405)	-0.894	0.561	0	0	
Semester	2	-0.628	0.449	0.202	(-1.508, 0.252)	-1.399	6.265*	1.077	84.0	
Assessment Setting							0.241			
Inclusive	2	-0.340	0.380	0.144	(-1.084, 0.405)	-0.894	0.561*	0	0	
Specialized	2	-0.628	0.449	0.202	(-1.508, 0.252)	-1.399	6.265*	1.077	84.0	
Assessment Design							0.301			
Descriptive	1	-0.206	0.555	0.308	(-1.294, 0.883)	-0.371	0	0	0	
Experimental	3	-0.566	0.350	0.123	(-1.252, 0.121)	-1.616	6.303*	0.241	68.2	
Sample Characteristics [b]										
Age (Grade Level)							1.112			
Elementary	2	-0.940	0.573	0.328	(-2.063, 0.183)	-1.641	3.668	0.496	72.7	
Middle	1	-0.034	0.767	0.589	(-1.583, 1.469)	-0.045	0	0	0	
Combined	1	-0.206	0.750	0.563	(-1.676, 1.264)	-0.274	0	0	0	
Country							6.850			
Australia	1	-0.034	0.305	0.093	(-0.632, 0.564)	-0.113	0	0	0	

	Effect Size Statistics					Null Test	Heterogeneity Statistics			Publication Bias
Canada	1	-1.635	0.562	0.316	(-2.737, -0.534)	-2.909	0	0	0	
S. Africa	1	-0.468	0.236	0.055	(-0.930, -0.006)	-1.987	0	0	0	
Study Characteristics [b]										
Measure							1.112			
Self-Report	1	-0.206	0.750	0.563	(1.676, 1.264)	-0.274	0	0	0	
Objective	1	-0.034	0.767	0.589	(-1.538, 1.469)	-0.045	0	0	0	

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 ($\alpha < 0.05$).

Discussion

The purpose of this meta-analysis was to collect an overview of research that provides effective means of assessment, in the affective domain of physical education, in adapted physical education settings. The results indicated an overall negative effect, with three of the four studies using an experimental design. In most cases, students with special needs were compared to a control group that consisted of their typically developing peers. Another factor to consider is that no study reported on more than one outcome, (self-concept, anxiety, language, social competence, depression, behavior, etc.), making it difficult to determine whether or not the assessments used were effective. A moderator analysis was conducted indicating possible factors that may have influenced the effectiveness of each intervention.

Assessment Approach

Summative assessment had a moderate, negative effect. 75% of the studies found used summative assessment as their assessment approach, making it difficult to determine student learning, and outcome, throughout the duration of the study. In a physical education setting, as in any academic setting, formative assessment is imperative to student success, and should be used to guide instruction as students are observed on an on-going basis. Formative assessment also serves the purpose of measuring and monitoring student learning throughout the lesson/assessment process. The evidence derived from formative assessment can be used to make decisions that best meet the needs of the students in adapted physical education. Without the use of formative

assessment, it is often unclear whether students are meeting standards and improving in any of the domains of physical education (Haug, 2015).

Assessment Duration

Semester studies had a moderate, negative effect, showing that when students are exposed to the same or similar material over time, their attention span is greater, making true student learning a more likely outcome. With 2 of the studies only being performed over the course of a unit, and two over the course of a semester, it is hard to determine whether or not the effect size would have been larger if the studies had taken place over longer periods of time. It can be hard to see change/progress in just a unit and/or semester. When interventions take place over the course of a year or longer, better decisions can be made when long-term effects are observed/assessed (Mercier & Lacovelli, 2014).

Assessment Setting

Specialized settings had a moderate, negative effect showing that students performed better when taught and assessed in a specialized setting. This can be due to slower paced instruction, smaller class size, a lesser student to teacher ratio, and differentiated instruction that is meant to meet the individual learning needs of all students (Hocutt, 1996). Two of the studies took place in inclusive settings, while the other 2 took place in specialized settings. Because it is hard to determine whether or not students underwent intervention and were assessed in a "least restrictive environment", more information is needed to determine if the assessment setting had any impact on the results of the intervention. Assessment/intervention results that are collected/observed in

a specialized setting cannot be generalized to similar results in an inclusive setting because environment can play a huge role on student behavior, retention, attention, performance, ability, etc (Hocutt, 1996).

Sample Characteristics

Age. Elementary age had a large, negative effect. This could be because elementary school is the first time students are exposed to any type of assessment. It could also be due to the fact that typically developing students have higher cognitive and emotional function (Valiente et. al. 2012). Two of the studies were conducted with elementary aged children, 1 with middle school aged children, and one with a combination of the two. No studies were conducted using high school aged students. While the data gained from interventions with younger children is beneficial in guiding evidence-based practice and assessment, intervention with older (high school aged) students would benefit these students soon going into adulthood. Especially in physical/adapted physical education settings, physical educators have the potential to teach students how to remain physically active throughout their lives, which has a huge impact on their health and overall affective/emotional existence (Kriemler et. al. 2011).

Country. Australia had a small negative effect while Canada had a large, negative effect. This can be caused by cultural differences in curriculum and learning focus as well as laws and regulations that facilitate students with disabilities' learning. These regulations can include resources provided such as teacher training, time for instruction, student learning materials available to students, etc. These effects could also be caused by different cultural contexts and views on children with disabilities. Smith (2014) stated

that “not all cultures share the same concepts of disability, thus disabilities must be viewed within a cultural context”. This is important to consider when examining the effects that different countries and cultures play when looking at student assessment and learning outcomes.

Measure. Objective measure had a small negative effect which shows a true representation of the results found. A combination of self-report and objective had a large, negative effect. This is probably because students with disabilities do not have the same cognitive abilities to be able to self-report and because their perception of reality may be skewed.

It was not specified in the 4 studies that were found, whether the students who underwent intervention were male or female. When looking at the affective outcome domain of physical education, gender can play a huge role on intervention/assessment outcomes. Gender is important when making decisions on what students gain when evidence-based assessments and teaching practices are used.

References

Bertelsen, S. (July 2002). An Analysis of the Relationships Between Teacher Behaviors and Affective Outcomes of Students in Middle School Physical Education. Retrieved from ProQuest. (3059975).

Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies, ISSN 0007-1005.* p108-121

Hansen, K. (2008). Teaching Within All Three Domains to Maximize Student Learning. *A Journal for Physical and Sport Educators,* 21, 9-13. https://eric.ed.gov/?id=EJ826300

Harlen W, Deakin Crick R (2002) A systematic review of the impact of summative assessment and tests on students' motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Haug, B., & Odegaard, M. (2015). Formative Assessment and Teachers' Sensitivity to Student Responses. *International Journal of Science Education.* v37 n4 p629-654 2015. 26 pp.

Hocutt, A. (1996). Effectiveness of Special Education: Is Placement the Critical Factor? *The Future of Children: Special Education for Students with Disabilities Vol. 6, No. 1.* p77-102.

Jin, J., & Yun, J. (2013, January 26). Evidence-based Practice in Adapted Physical Education. *Journal of Physical Education, Recreation & Dance,* 81:4, 50-54, DOI: 10.1080/07303084.2010.10598465

Jones, C. A. (2005). Assessment for Learning. Learning and Skills Development Agency. Retrieved from http://dera.ioe.ac.uk/7800/1/AssessmentforLearning.pdf.

Kelly, L. E. (Eds.) (2006). Adapted Physical Education National Standards. Second Edition. United States: National Consortium for Physical Education and Recreation for Individuals with Disabilities.

Kriemler S, Meyer U, Martin E, *et al*. (2011). Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update. *British Journal of Sports Medicine*;**45:**923-930

Lahey, J. (2014, January 21). Students Should be Tested More, Not Less. *The Atlantic.* Retrieved from https://www.theatlantic.com/education/archive/2014/01/students-should-be-tested-more-not-less/283195/

Mercier, K., & Lacovelli, T. (2014, January 30). Summative Assessments: How we Improved our High School Physical Education Program. *Journal of Physical Education, Recreation & Dance.* p14-18.

NASPE (2011). Retrieved from https://pgpedia.com/n/national-association-sport-and-physical-education

Nicol, D. J., & Macfarlane-Dick, D. (2007). Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education, 31:2,* 199-218. doi: 10.1080/03075070600572090.

Peens, A., Pienaar, A.E., & Nienaber, A.W. (2007, September 3). The Effect of Different
Intervention Programmes on the Self-oncept and Motor Proficiency of 7- to 9-year-old children with DCD. *Child Care Health and Development.* p316-328.

Smith, D. D. (2007). Differing Perspectives on Disabilities. *Special Education: Making a Difference.* p7-9.

Starline, J., Munro, N., Togher, L., & Arciuli, J. (2012). Training Secondary School Teachers in Instructional Language Modification Techniques to Support Adolescents with Language Impairment: A Randomized Controlled Trial. *Clinical Forum: Language and Communication Disorders in Adolescents.* p474-495.

Todorovich, J. (2013, July 26). Evidence-based Physical Education #1: What Strategies Can We Use to Increase Students' MVPA in Physical Education Lessons? *Journal of Physical Education, Recreation & Dance.*

Valiente, C., Swanson, J., & Eisenberg, N. (2012). Linking Students' Emotions and Academic Achievement: When and Why Emotions Matter. *US National Library of Medicine National Institutes of Health.* p129-135.

Verret, C., Guay, M., Berthiaume, C., Gardiner, P., & Beliveau, L. (2012). A Physical Activity Program Improves Behavior and Cognitive Functions in Children with ADHD: An Exploratory Study. *Journal of Attention Disorders.* p71-80.