

DEVELOPMENT OF A RESEARCH TAXONOMY FOR
ADAPTED PHYSICAL ACTIVITY

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE GRADUATE SCHOOL OF THE
TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF KINESIOLOGY
COLLEGE OF HEALTH SCIENCES

BY
SHARON L. CARANO, MS, CAPE

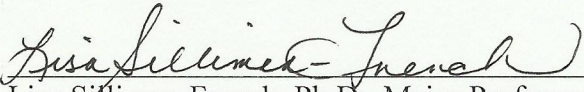
DENTON, TX
DECEMBER, 2014

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS

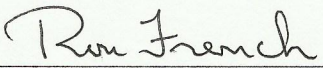
November 11, 2014

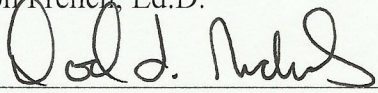
To the Dean of the Graduate School:

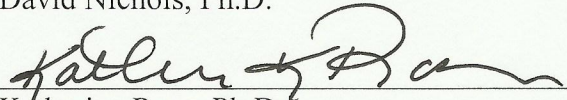
I am submitting herewith a dissertation written by Sharon L. Carano entitled "Development of a Research Taxonomy for Adapted Physical Activity." I have examined this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Kinesiology.



Lisa Silliman-French, Ph.D., Major Professor

We have read this dissertation and recommend its acceptance:

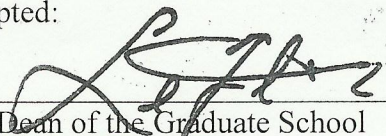

Ron French, Ed.D.


David Nichols, Ph.D.


Katherine Rose, Ph.D.


Department Chair

Accepted:


Dean of the Graduate School

ACKNOWLEDGMENTS

The end of this very long journey has come to a close with a new one beginning. This journey would not have been possible without the support and encouragement of many people. God has blessed me and surrounded me with a great group of champions.

First, I would like to thank Dr. Silliman-French for agreeing to join me on this half-baked idea. To Dr. French for giving us what seemed to be a rather benign assignment that gave birth to this half-baked idea? To you both thank you for sticking it out and for Dr. Nichols and Dr. Rose for joining us on this journey.

To my family, God did something funny when he took this DC girl and landed her in Texas, who would have thought? Mom and Dad I am indebted to you for constant support, the trips you made, and everything else you did. It all seemed to come together at just the right time.

To my husband and Paz, you both have been a constant and steady support, to know I could always trust in your guidance when my vision was blurry. I may walk away from this journey with a PhD but I count it greater to have encountered you both during this time.

Finally, to my friends around the country who at varying seasons of my life supported me from afar, I say thank you. To my girls in Maryland; the decades have passed, the miles have grown longer, but no matter how tough things would get I knew I had an army encouraging me.

ABSTRACT

The purpose of this study was fourfold to develop an Adapted Physical Activity Taxonomy (APAT). Purpose one was to determine through a systematic literature search to identify research taxonomies. Purpose two was to identify categories and criteria present within published research taxonomies. Based on the findings of the literature search, purpose three was to identify what categories and criteria are needed in the APAT. Purpose four was to determine the inter-rater reliability (IRR) of the APAT. A modified mixed method process was used in a meta-framework design (Onwuegbuzie et al., 2010).

After an in depth literature search this is the first research investigation designed to develop an Adapted Physical Activity Taxonomy. An expert panel participated in the Delphi Survey to identify the categories and criteria included in the APAT. The present investigator could not determine IRR because the return rate was insufficient for statistical analysis; therefore, the findings were inconclusive.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
 Chapter	
I. INTRODUCTION	1
Adapted Physical Activity and Research Evaluation	2
Increasing the Quality of Research in Adapted Physical Activity	5
Evidence-based Practice and Research Taxonomies	6
Evidence-based Practice in Education	8
Summary	9
Purpose	9
Research Questions	10
Definitions	10
Limitations	12
Delimitations	12
II. REVIEW OF LITERATURE	13
State of Research in Adapted Physical Activity	14
Level of Evidence	14
Strength of Recommendation	15
Evidence-based Practice	23
Role of Evidence-based Practice in Education	23
Council for Exceptional Children and Evidence-based Practice	24
History of Taxonomies/Grading Systems	25
Users of Taxonomies	25
Purpose of Taxonomies	26
Evolution of Taxonomies	27
Use of Taxonomies	28

Development and Use of Two Comprehensive Taxonomies	36
Grading of Recommendations, Assessment, Development, and Evaluation ..	37
Strength of Recommendation Taxonomy	40
Example of Two Prominent Taxonomies	44
Council for Exceptional Children Quality Indicators	47
Correlational	48
Experimental/Quasi-Experimental	50
Qualitative	51
Single-Subject	52
 III. METHOD	 56
General Research Design	57
Study 1: Phase I – Literature Analysis	58
Study 2: Phase II-IV	60
Phase II: Delphi Survey Technique	60
Phase III: Adapted Physical Activity Taxonomy Development	63
Phase IV: Pilot Study	63
Study 3: Phase V	64
Inter-Rater Reliability	64
 IV. RESULTS	 66
Study 1: Literature Analysis - Phase I	67
Literature Analysis	68
Study 2: Phases II-IV	72
Delphi Survey Technique Phase II	72
Round 1	73
Round 2	77
Round 3	78
Adapted Physical Activity Taxonomy Development Phase III	79
Pilot Study Phase IV	83
Study 3: Phase V	85
Inter-Rater Reliability	85
Correlational Design	87
Experimental/Quasi-Experimental Design	88
Qualitative Design	88
Single-Subject Design	89

V. DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER STUDIES	91
Discussion	91
Systematic Literature Analysis	93
Identifying Categories and Criteria	95
Determine Criteria Needed for Adapted Physical Activity Taxonomy	96
Determine Inter-Rater Reliability	98
Conclusions	98
Recommendations for Further Studies	99
REFERENCES	101
APPENDICES	
A. Strength of Recommendation Taxonomy Summary of Adapted Physical Activity Articles	116
B. First Round of Delphi Survey Documents and Questions	120
C. Second Round of Delphi Survey Questions	126
D. Third Round of Delphi Survey Questions	130
E. Adapted Physical Activity Taxonomy Instruction Sheet	132
F. Final Adapted Physical Activity Taxonomy: Correlational Design	134
G. Final Adapted Physical Activity Taxonomy: Experimental/ Quasi-Experimental Design	139
G. Final Adapted Physical Activity Taxonomy: Qualitative Design	144
I. Final Adapted Physical Activity Taxonomy: Single-Subject Design	149

LIST OF TABLES

Tables	Page
1. How to Evaluate Whether an Educational Intervention is Supported by Rigorous Evidence: An Overview	24
2. A Summary of Grading of Recommendations, Assessment, Development, and Evaluation's Approach to Rating Quality of Evidence Rating Symbol	39
3. Strength of Recommendation Rating Symbol	40
4. Assessing Quality of Evidence with SORT	42
5. Strength of Recommendation Grades	43
6. Example of Quality of Study and Recommendation of SORT and GRADE Taxonomies	45
7. Database and Numbers of Located Articles	68
8. Number of Research Designs Evaluated	71
9. Participant Demographics	74
10. Nvivo Analysis of Correlation Coefficient for Participants Responses	77
11. Sample of Suggested Participant Changes	81
12. Sample of Summary of Participant Suggestions	82
13. Participant Recruitment	85
14. Summary of Taxonomy Development Findings	94

LIST OF FIGURES

Figures	Page
1. Example of Dissertation use of a Taxonomy	30
2. Example of Meta-analysis use of a Taxonomy	32
3. Example of a Taxonomy Incorporating Findings Within a Chart	33
4. Example of use of a Taxonomy in the Method Section of Literature Reviews within Experimental Research	35
5. Quality Indicators for Correlational Research	49
6. Essential and Desirable Quality Indicators for Group Experimental and Quasi-Experimental Research Articles and Reports	51
7. Quality Indicators Within Qualitative Research	52
8. Quality Indicators Within Single-Subject Research	54
9. Study and Phases	57
10. Instrument Development and Construct Validation Process	58
11. Visual Representation of Purposes, Three Studies, and Five Phases of the Investigation	67
12. Percent of Participants Indicating Strong, Moderate, or Weak SOQ for Correlational Research Design Article	87
13. Percent of Participants Indicating Level A, B, or C LOR for the Correlational Research Design Article	87
14. Percent of Participants Indicating Strong, Moderate, or Weak SOQ for Experimental/Quasi Experimental Research Design Article	88

15. Percent of Participants Indicating Level A, B, or C LOR on for the Experimental/Quasi Experimental Research Design Article	88
16. Percent of Participants Indicating Strong, Moderate, or Weak SOQ for Qualitative Research Design Article	89
17. Percent of Participants Indicating Level A, B, or C LOR for Qualitative the Research Design Article	89
18. Percent of Participants Indicating Strong, Moderate, or Weak SOQ for Single-subject Research Design Article	90
19. Percent of Participants Indicating Level A, B, or C LOR for the Single-Subject Research Design Article	90

CHAPTER I

INTRODUCTION

For over 28 years, researchers in the profession of Adapted Physical Activity have stated that current research lacks quality (Lavay & Lasko-McCarthy, 1992) and there is a need to develop more rigorous methods for conducting and evaluating the quality of research produced (Bouffard, 1997). One possible solution is the use of a research taxonomy that determines the quality of evidenced-based practices. Research taxonomies are used in many professions and professions to substantiate the quality of prior research (e.g., pediatrics, nursing, speech and language, physical therapy). Users of research taxonomies include patients, graduate students, professors, and/or reviewers of journal articles; and it varies how each evaluates the quality of research. While some research taxonomies are used to evaluate individual study designs, other research taxonomies are designed to incorporate multiple designs (e.g., randomized control trial, mixed method, single-subject). Taxonomies identify the quality of the study and to what extent the treatment or protocol should be recommended, also known as the level of recommendation.

Research taxonomies can also be applied within a literature review of a study in the method section. By indicating the quality of a research article, taxonomies can provide an evaluation criterion supporting the use of specific research within the literature review. When a researcher can determine the quality of prior studies that are included in a

literature review, he/she can substantiate the level of quality for inclusion and justification for citing that research. Many taxonomies are used to identify key categories (i.e., major headings in research articles), and criteria (i.e., key statements) of importance, which are used to evaluate and produce a score. These evaluation score can then be aligned within a specific profession. These scores may be known as quality of study and level of recommendation.

Presently, there have been no research taxonomies identified through an extensive literature search that clearly align with the profession of Adapted Physical Activity. Therefore, research taxonomies have been used from other professions to evaluate the quality of research conducted in the profession. As in other professions, the development of a research taxonomy for Adapted Physical Activity could raise the quality of research and hold researchers more accountable. It could also address the concerns of the Adapted Physical Activity community for stronger and higher quality research.

Adapted Physical Activity and Research Evaluation

The term Adapted Physical Activity is less than 50 years old, and was first introduced in 1973 by founders of the Federation Internationale de l' Activite Physique Adaptee (Hutzler & Sherrill, 2007). The term Adapted Physical Activity serves as an umbrella term (DePauw & Doll-Tepper, 1989) to include the professions of adapted physical education (APE); adapted recreation, known as therapeutic recreation; and adapted sport, also known as disability sport (Porretta, Nesbitt, & Labanowich, 1993).

As the profession of Adapted Physical Activity continues to develop into a profession (O'Connor, French, Sherrill, & Babcock, 2001; Sherrill & O'Conner, 1999), there has been a demand to improve and evaluate the quality of Adapted Physical Activity research (Jansma & Surburg, 1995). Evaluating the quality of research in Adapted Physical Activity holds researchers accountable for conducting quality research and results in recommendations for researchers and practitioners. These recommendations will further support and substantiate the profession of Adapted Physical Activity as a profession (Reid & Stanish, 2003).

More than 406 articles between 1984 and 2011 have been published in *Adapted Physical Activity Quarterly (APAQ)* regarding research related to the profession of Adapted Physical Activity (Porretta & Sherrill, 2005; Reid & Stanish, 2003). Some of these articles included information pertaining to the quality of the research publications in *APAQ*. As research has developed in Adapted Physical Activity, a greater emphasis has been placed on the need for: (a) theory to drive research, (b) theory to substantiate research (Reid, 1989), and (c) exploration of past research in an effort to develop higher quality research (Bouffard, 1997). Furthermore, Porretta, and Sherrill (2005) identified *APAQ* articles published between 1994 and 2003 that related to the topic of research and stated that research related articles were defined as “. . . any topic related to research methods or practice . . .” (p. 122).

In 1992, Lavay and Lasko-McCarthy presented issues and recommendations in the profession of Adapted Physical Activity related to increasing the quality of research, as

well as, the need to consider alternative research designs that better suit the profession of Adapted Physical Activity. One aspect of producing research that can positively impact the profession of Adapted Physical Activity is to focus on relevant research. Reid (1993) stated that newly developing Adapted Physical Activity research should “. . . become relevant to instruction if a research program is created that increasingly moves toward applied questions” (p. 361). To facilitate this process, Reid further stated that quality research be designed to: (a) address conceptual issues, (b) provide other researchers enough information for replication, (c) implement quality statistical methods, and (d) be generalizable results outside a controlled setting.

In an effort to address Adapted Physical Activity research quality, Jansma and Surburg (1995) developed and published competency guidelines for Adapted Physical Activity doctoral candidates. The Adapted Physical Activity doctoral competencies were intended to prepare quality researchers, thereby enhancing the quality of published research in the future in the profession of Adapted Physical Activity. For instance, three doctoral competencies listed by the authors were:

1. A researcher will “Select, apply, and interpret statistics appropriate for selected research designs” (p. 312).
2. A researcher will “Write the results of a research problem in publishable form . . .” (p. 312).
3. A researcher will “Maintain an awareness of relevant issues and trends in research” (p. 313).

With the development of the Adapted Physical Activity Taxonomy (APAT) these competencies can be addressed.

Increasing the Quality of Research in Adapted Physical Activity

Several authors have addressed ways to increase the quality of research in the profession of Adapted Physical Activity. Two key factors identified were the need to: (a) provide a full description of the instruments used (Sherrill, 1997); and (b) describe participants clearly enough that other researchers would be able to replicate the population (Reid, 1993). Sherrill and O'Connor (1999) continued to address the need to improve the quality of research in Adapted Physical Activity. Some of their criteria were:

- to present a theoretical foundation or conceptual framework;
- to clearly state methods with the statement of hypotheses or research questions;
- to provide a full description of participants;
- to clearly identify instruments; and
- to provide the validity and reliability of the instrument(s) used.

Sherrill and O'Connor also suggested the use of the *Publication Manual of the American Psychological Association* to identify generally accepted areas to report on within published research and to address key components that related specifically to Adapted Physical Activity. Areas of improvement noted were, the identification of key areas that were to be addressed, and what should be included and published as quality research (Porretta & Sherrill, 2005; Sherrill & O'Connor, 1999). The various viewpoints

detailing the characteristics of quality research in Adapted Physical Activity supports the argument for a taxonomy to evaluate the quality of research in the profession of Adapted Physical Activity (see also, Belanger & Colantonio, 2011).

Evidence-based Practice and Research Taxonomies

Evidence-based practice (EBP) has become a driving force behind what is considered quality research. Developed within the medical community, EBP is also known as evidence-based medicine (Bouffard & Reid, 2012). Portney and Watkins (2007) defined EBP as a “. . . necessity for documenting and testing elements of clinical practice through rigorous and objective analysis and scientific inquiry . . .” (p. 3). Since the need for EBP has increased, questions on how to best determine the specific quality of research has arisen. One possible answer to this question was in the form of what has become known as a taxonomy.

In order to most effectively evaluate and substantiate EBP, published research must be evaluated by a set of evidence-based criteria. Cochrane (1972) has been credited with the development of an EBP instrument to evaluate research based on an identified need. This instrument was designed to improve research quality and to identify quality research, also known as a research taxonomy.

Based on a study by Ebell et al. (2004), more than 140 research taxonomies have been developed in various professions from 1979 to 2004. Since this study, other experts in other professions have developed research taxonomies including pediatrics (Marcuse & Shiffman, 2004), speech and language (Gillam & Laing, 2006), nutrition (Woolf,

2006), dentistry (Faggion, 2010), music therapy (Abrams, 2010), and nursing (Pilcher & Bedford, 2011). The majority of these research taxonomies are in the health professions; however, professionals in other professions have also developed research taxonomies to meet the uniqueness of their specified areas of research. There are currently no known research taxonomies in the profession of Adapted Physical Activity, thus leaving the profession to use other disciplines for guidance.

In 1979, the first research taxonomy for evaluating published research was developed by a Canadian Task Force with a primary emphasis on randomized control trial (RCT) studies. While the RCT is, and should be considered the highest quality of research design, it does not negate the credibility of other designs (Bouffard & Reid, 2012; Zhang, deLisle, & Chen, 2006). In the last 10 years, research taxonomies in the medical profession have included research designs other than RCT (e.g., single-subject designs).

Many current research taxonomies are used first to evaluate the quality of a study followed by recommendations of the study for practice (Ebell et al., 2004). Research taxonomies range from a word description for the evaluation of a study, or a numerical evaluation such as 1, 2, or 3 to represent an evaluation score. It is important to note that a study can be considered high quality and not be recommended for practice. An example might be the quality of the study is high but the intervention or treatment might be too limited (Ebell et al.).

Therefore, in addition to the quality of a specific study, the second component of research taxonomies are the recommendations of the EBP. The recommendation of a

study is provided to assist the reader in understanding the extent a treatment or intervention (i.e., highly recommend, moderately recommended, limited recommendation, not recommended) should be implemented. The primary aim of the medical community's research taxonomies relates to the treatment of disease. This limitation is what brought about the development of research taxonomies for other professions. The aim of the Adapted Physical Activity profession is not directed at disease treatment and therefore, there is a need for an Adapted Physical Activity taxonomy.

Evidence-based Practice in Education

The enactment of the No Child Left Behind Act (NCLB, 1994), an act aimed to improve the quality of education for students in the United States, established an increased demand for EBP within the profession of education. The term "scientifically based research" in NCLB is another term for EBP (Allington, 2006). In the profession of special education, there has been an interest for higher quality research and the means to substantiate EBP. A set of indicators for determining quality of EBP in research has been published (Institute of Education Sciences, 2003). Alternatively, the Council for Exceptional Children (CEC, 2004) has published criteria that identified strong quality indicators for various research designs. These criteria were used to identify quality indicators for specific research designs that provide guidance for researchers in the profession of special education and can be applied to Adapted Physical Activity profession.

While it has been noted that there are numerous research taxonomies, to date there are no known research taxonomies that could be located that evaluate and provide an evaluation of the quality of research and its contribution to the existing body of literature in the profession of education (Stuart, Tondora, & Hodge, 2004); specifically in special education (Odom, Brantlinger, Gersten, Horner, Thompson, & Harris, 2005); or Adapted Physical Activity (Belanger & Colantonio, 2011).

Summary

In summary, an Adapted Physical Activity Taxonomy (APAT) is needed to determine the quality of research and its contribution to EBP. The taxonomy will be used to substantiate practices based on interventions, as well as, identify the quality of research. These interventions should be able to withstand scrutiny of what is considered to be quality within Adapted Physical Activity literature (Levin, 2004). The development of an APAT must be as unbiased as possible, and designed to evaluate research based on key components as established by the professional community. It is the aim of the researcher to develop research taxonomy for the profession of Adapted Physical Activity.

Purpose

The purposes of this research project are fourfold related to the development of an APAT. The first purpose of this investigation was to determine through a systematic search of literature what research taxonomies exist in the English language, are used in any profession, and are embedded within published literature. A second purpose was to identify categories and criteria present within the published research taxonomies. Based

on the findings of the literature search, the third purpose was to identify categories within the research articles and what criteria are needed in the APAT. The fourth purpose was to determine the inter-rater reliability (IRR) of the categories and criteria included in the APAT that were developed.

Research Questions

Research is driven by asking questions (Thomas, Nelson, & Silverman, 2005). Those used to guide the current study were:

1. What are there essential criteria that need to be included in an Adapted Physical Activity taxonomy?
2. What categories and criteria can be determined through the use of a validation process?
3. What level of inter-rater reliability of an Adapted Physical Activity taxonomy can be established?

Definitions

Adapted Physical Activity: “The cross-disciplinary body of knowledge directed toward the identification and solution of individual differences in physical activity. It is a service delivery profession and an academic profession of study that supports an attitude of acceptance of individual differences, advocates access to activity lifestyles and sport, and promotes innovation and cooperative service delivery programs and empowerment systems. Adapted Physical Activity includes, but is not limited to, physical education, sport recreation, dance and creative arts, nutrition, medicine, and rehabilitation”

(International Federation of, retrieved 2012). In addition this term encompasses the following professions and professional areas (Porretta, Nesbitt, & Labanowich, 1993):

1. Adapted physical education: “Individuals who have been identified by schools or agencies as having handicapping conditions as well as those persons who have not been so formally labeled (p. 91).”
2. Adapted recreation: “Modified recreational activities provided both in treatment and community settings, facilitating participation in recreational activities on the part of disabled individuals (p. 92).”
3. Adapted/disability sport: “Participation in regional, national, and international competition under the governance of interscholastic, intercollegiate, and professional affiliations as well as formalized sport organizations (p. 93).”

Evidence-based practice: “The strategy or intervention designed for use by special educators and intended to support the education of individuals with exceptional learning needs” (Classifying the State of Evidence, 2008, p. 6).

Researcher: “The pursuit of lifetime of inquiry and ongoing programs of research which address the need to search for new answers, to critique existing practices, and to explore problems of interest to the researcher” (Jansma & Surburg, 1995, p. 312).

Research taxonomy: “The quality of evidence indicates the extent which we can be confident that an estimate of effect is correct; the strength of a recommendation indicates the extent to which we can be confident that adherence to the recommendation will do more good than harm” (Oxman, 2004, p. 1490). A working definition of research

taxonomies for this study is to evaluate the quality of research and provide an evaluation score to represent the level of quality.

Limitations

Because there are restrictions based on the design and the population used in this investigation, there are possible limitations that the researcher could not control and are as follows.

1. Research taxonomies are fairly new, originating in 1979 by the Canadian Task Force, therefore there are no consistent references to describe the various terms for classifying research taxonomies (e.g., research taxonomies).
2. Since English may not be the primary language of some participants, there may be differences in interpretation and terminology of key words and concepts.
3. May be the personal bias of participants because of their possible employment requirements that may negatively influence attitudes to evaluate and critique during the developmental process.
4. Possible research bias of the participants on the development of the research taxonomy instrument.

Delimitations

To control external influences that are not in the scope of this investigation, the researcher has imposed certain restrictions:

1. Confining the participants to those in the profession of Adapted Physical Activity.
2. Determining the research designs of most importance to the participants.

CHAPTER II

REVIEW OF LITERATURE

The purposes of this research project were fourfold related to the development of an Adapted Physical Activity Taxonomy (APAT). The first purpose of this investigation was to determine through a systematic search of literature what research taxonomies existed in the English language, are used in any profession, and are embedded within published literature. A second purpose was to identify categories and criteria present within the published research taxonomies. Based on the findings of the literature search, the third purpose was to identify categories within the research articles and what criteria were needed in the APAT. The fourth purpose was to identify and to validate the categories and criteria that were developed. Because a taxonomy for the profession of Adapted Physical Activity has not been developed for the evaluation of research, other professions with taxonomies were investigated. These taxonomies were studied for commonalities within the different taxonomies that needed to be considered for the development of an APAT.

The information in this investigation will answer the following question: “What is the state of research in Adapted Physical Activity?” A taxonomy is applied to the literature within this section to determine the level of evidence and the strength of recommendation. Following the presentation of this information, the following topics were explored: evidence-based practice (EBP); the role of EBP in education; the role of

the Council for Exceptional Children (CEC) and EBP; the history of taxonomies/grading systems, users of taxonomies, the purposes of taxonomies, the evolution of taxonomies, the use of taxonomies; and development and use of two comprehensive taxonomies, with accompanying examples of the two most published taxonomies, and the CEC quality indicators.

State of Research in Adapted Physical Activity

The existing literature was searched in order to determine the body of literature relevant to the profession of Adapted Physical Activity and research. Throughout the course of this literature review the Strength of Recommendation Taxonomy (SORT) was applied to determine the quality of study and recommendation of the research (Ebell et al., 2004). Based on SORT, the body of evidence is given a rating and the quality of the study ranked. Each of the articles was identified with a grade and a review of the article can be located in Appendix A. Next are key terms necessary for consistency and clarification and have been outline by Ebell et al.

Level of Evidence

Levels of evidence can be twofold and refer not only to an individual study but also to the quality of evidence from multiple studies about a specific question or the quality of evidence supporting a clinical intervention (Newman, Weyant, & Hujoel, 2007). Within this chapter and after the citation of the author(s), the level of evidence will be provided as: (a) L1 for a Level 1 study; (b) L2 for a Level 2 study; and (c) L3 for a Level 3 study. Newman et al., (p. 149) identified the three levels as follows:

- L1 study consists of “good-quality, patient-oriented evidence”
- L2 study consists of “limited-quality patient-oriented evidence”
- L3 study consists of “other evidence”

Strength of Recommendation

The strength of a recommendation for clinical practice is based on a level of evidence (Newman et al., 2007) and will appear within the citation of an article as an “A” indicating good evidence, “B” indicating inconsistent or limited evidence, and “C” indicating evidenced based on a lower level of study design (i.e., consensus, practice, opinion, case series, treatment, prevention). However, if a study is not about diagnosis or treatment, it cannot be assigned a strength of recommendation evaluation score.

The profession of Adapted Physical Activity is a fairly young profession, with the first recorded programs for individuals with disability noted in the 1950s (as cited in Sherrill, 2004). After the 1950s, the term Adapted Physical Activity gained support as an umbrella term for the profession (Broadhead, 1990, L3; DePaw & Doll-Tepper, 1989, L3). The profession of Adapted Physical Activity, which continues to be substantiated as a valuable profession, (O’Connor, French, Sherrill, & Babcock, 2001, L3; Sherrill, 1997, L3) encompasses the professions of adapted physical education (APE), disability sport, and adapted recreation. Early research in the profession of Adapted Physical Activity was thought to be outcome driven, focused on the performance of individuals with disabilities, was disability specific (i.e., blind, intellectual disability), and focused primarily on key elements of motor development (Doll-Tepper & DePaw, 1996, L3).

Other lines of developed research have focused on the use of surveys (Porretta, Kozub, & Lisboa, 2000, L3) and the inclusion of individuals with disabilities (Doll-Tepper & DePauw, 1996, L3).

In 1984, researchers in the profession expressed concern over the need to identify how research has been conducted, as well as, the need to identify future lines of research based on what has been accomplished in the past (Reynolds, 1984, L3). Of particular interest in the 1990s was the need to link research with known theoretical practice. It was suggested that by grounding research in theory it would provide a solid foundation for conducting research and develop a clear line of future research (Reid, 1993, L3).

Another approach for identifying strong research was presented in the 90s was Bouffard's (1993, L3) who suggested the use of person-by-treatment interaction (PTI) in the profession of Adapted Physical Activity. Bouffard strongly believed in the need for a shift from less quantitative research to more qualitative research and posed an argument about the pitfalls in attempting to categorize such a diverse population within quantitative research. Bouffard's (1993, L3) argument was that researchers should focus first on the person followed by the treatment (Shephard, 1999, L3).

O'Connor et al. (2001, L3) identified, through bibliometric research, scholarly productivity in the profession of Adapted Physical Activity and confirmed that scholarly productivity exists within the profession. Over the last two decades the profession of Adapted Physical Activity has continued to grow and researchers have become more critical of the quality of research published in the profession (Porretta & Sherrill,

2005, L3). Porretta and Sherrill (2005) were able to identify more than 400 articles related to the topic of research published over a 20-year period. The compilation of these articles addressed issues related to improving the quality of research (Sherrill & O'Connor, 1999, L3), improving research methods, and the importance of statistical analysis (Sutlive & Ulrich, 1998, L3).

In the pursuit of higher quality research, researchers have identified key areas of improvement that are needed in the profession of Adapted Physical Activity. It is thought, regardless of what research model is followed (Bouffard, 1997, L3), which theoretical prospective is identified (Sherrill, 1997, L3), or the philosophical foundation that is cited, (Bouffard, Strean & Davis, 1998, L3; Reid, 2000, L3) that the quality of research must increase (Doll-Tepper & DePauw, 1996, L3).

Broadhead and Burton (1996, L3) stated that in order for the profession of Adapted Physical Activity to continue to be recognized as having a strong knowledge base, researchers must continue to produce quality research. One aspect of developing quality research addressed by Strean (1998, L3) was the ability of researchers to identify assumptions related to research. Strean suggested that a reader can only fully understand the perspective of the researcher by identifying and drawing into question assumptions of research. Assumptions play a critical role in the development of a study to ensure that all possible biases, known and unknown, have been addressed (Strean, 1998, L3).

Reid and Stanish (2003, L3) eluded that the profession of Adapted Physical Activity is linked to the medical profession and the profession of education because of the strong

correlation in terminology (Broadhead & Burton, 1996, L3; as cited in Sherrill, 2004, L3). Reid and Stanish (2003, L3) suggested that the profession of Adapted Physical Activity is a cross discipline rather than its own independent profession. With a relationship between the two, it is no surprise the profession of Adapted Physical Activity would refer to the medical profession and how it has developed a means for evaluating EBP in an effort to substantiate the quality of evidence of research. The profession of Adapted Physical Activity has also begun turning to the profession of education, specifically special education, for indicators for evidence-based practice (Kassner, Reid, & MacDonald, 2012, L3).

In an effort to categorize research in APA, Christina (1989, L3) published the first known standards on classifying motor learning research by categorizing research into three levels of research based on application. These levels are:

- Level 1 – Least direct relevance, Basic research
- Level 2 – Moderate direct relevance, Applied research
- Level 3 – Most direct relevance, Applied research

The levels, as described by Reid (1989, L3), suggest that each category of research is independent and each is relevant. Rather than identifying the quality of research based on EBP, the levels identify research quality based on if the research is, or is not, theory driven.

Following Christina's publication, Lavay and Lasko-McCarthy (1992, L3) identified four potential issues on conducting quality research. The issues identified were;

“... difficulties in acquiring large and homogenous samples; developing valid, reliable, and commercially available test instruments and protocols specific to persons with disabilities; training doctoral students to conduct quality research; and maintaining a specific research focus” (p. 189-190, L3). These were identified in an effort to encourage higher quality research.

In an effort to improve the quality of future research, Adapted Physical Activity doctoral competencies were developed by Jansma and Surburg (1995, L3). These competences were dedicated to developing future doctoral students as quality researchers. A study measuring the attainment of these competences of doctoral students was conducted by Porretta, Surburg, and Jansma (2002, L3). Results indicated an increase in the confidence of doctoral students in attaining research competences. It was speculated this was a result of concern by professors for the need to increase doctoral student's aptitude in conducting quality research (Porretta et al., L3).

Following the publication of these competences several articles were published related to the topic of quality research. Sutlive and Ulrich (1998, L3) outlined six recommendations based on a review of literature for developing and reporting statistically significant research:

- To calculate and report effect size. By calculating and reporting effect size it informs the reader of the degree of differences between groups being measured (American Psychological Association, 2010).

- To consider selecting a larger alpha in order to determine statistical significance. By setting a larger alpha this can decrease the likelihood of committing a Type II error.
- To evaluate results within a sample size context thereby considering three variables in conjunction with one another: effect size, alpha level, and power in order to determine sample size.
- To emphasize replication of results. This can be done when the researcher provides a substantial amount of information to the reader so that he/she is able to conduct an independent study of the same nature with a different group of participants.
- To use “simple” research designs that minimize the number of dependent and independent variables.
- To use clear modifiers in describing “significance.” The term statistically significant should be considered for use when one rejects the null hypothesis. This indicates to the reader that the researcher has given consideration to the statistical meaningfulness of the results.

Following Sutlive and Ulrich (1998, L3), Sherrill and O’Connor published “Guidelines for Improving Adapted Physical Activity Research” (1999, L3). The following are the guidelines proposed:

- Strive for theoretical richness and stimulates exchange (Reid, 1992, L3; Sherrill, 1997, L3);

- Use hypotheses or research questions (American Psychological Association, 2010);
- State sampling design and discuss generalization of findings (Bouffard, 1993, L3);
- Describe participants fully;
- Avoid combining dissimilar subgroups;
- Use person first terminology and avoid bias in language (American Psychological Association, 2010);
- Describe sample-specific validity and reliability for each instrument;
- Report means, standard deviations, and effect sizes (American Psychological Association, 2010; Sutlive & Ulrich, L3); and
- Provide sufficient information for replication (Sutlive & Ulrich, 1998, L3).

Further research conducted by Zhang, deLisle, and Chen (2006, L3) indicated a growing number of publications of non-experimental design. Literature in the profession of Adapted Physical Activity identifying what would be considered to be strong qualitative research was not located. Zhang et al. were able to determine that research abstracts related to special populations most commonly used a group, time series, ex post facto, or descriptive research design (2006, L3). This suggests the need to include multiple types of research design within an APAT.

More recently Reid, Bouffard, and MacDonald (2012, L3) addressed the need to develop evidence-based research in Adapted Physical Activity. Reid et al. identified five

beliefs, which are substantiated by strong research and must be considered for the development of EBP. These five beliefs are: “individualization, critical thinking, self-determination, program effectiveness, and multifactor complexity” (Bouffard, & MacDonald, 2012, p 118). Furthermore, Bouffard and Reid (2012, L3) acknowledge the growing trend in the use of EBP in the profession of Adapted Physical Activity and outline 10 questions that should be asked to lead to better practices. As outlined by Bouffard and Reid these 10 questions were:

1. What is evidence-based practice?
2. What counts as evidence?
3. What are the rationales behind the different hierarchies of evidence?
4. How do we generalize from aggregate to person?
5. Are most research results universal or contextual?
6. Should all decisions about what works be evidence-based?
7. Why is evidence-based practice not evidence-based?
8. Should we endorse current governing practices?
9. Shall the domain of acceptable inquiry be restricted?
10. Will evidence-based practice work?

The use of research taxonomy addresses the growing concern to support EBP. Hutzler (2012, L3) identified that EBP is the driving force behind quality research, while others have used key quality indicators adopted from the profession of special education to address levels of evidence (Hutzler, L3; Kassner et al., 2012, L3). A survey of interest

and knowledge about the use of taxonomy was conducted by Belanger and Colantonio (2012, L3) that indicated a need for the development of an evidence-based research taxonomy for the profession of Adapted Physical Activity.

Evidence-based Practice

The term EBP has been derived from the medical profession and is defined as “. . . an approach to healthcare wherein health professionals use the best evidence possible, i.e., the most appropriate information available, to make clinical decisions for individual patients” (McKibbin, 1998, p. 396). Federal legislation such as No Child Left Behind (NCLB, 2004) has called for the use of scientifically based research in an effort to substantiate and to ensure that sound teaching practices and methodologies were being used in the classroom. The CEC defined EBP for special education as a “. . . strategy or intervention designed for use by special educators and intended to support the education of individuals with exceptional learning needs” (Classifying the State, 2008, p. 6).

Role of Evidence-based Practice in Education

According to the United States government, the role of evidence-based practice is to better identify and use EBP to implement sound educational practice (Institute of Education Sciences, 2003). The government provides a three-step process that can be used to evaluate whether an intervention is based on, and supported by rigorous evidence. A modified version of these steps is presented in Table 1. Within this process an EBP can only earn a level of strong if the study design is a randomized control trial (RCT).

Table 1

How to Evaluate Whether an Educational Intervention is Supported by Rigorous Evidence: An Overview

Step 1	Is the intervention backed by “strong” evidence of effectiveness?
Step 2	If the intervention is not backed by “strong” evidence, is it backed by “possible” evidence of effectiveness?
Step 3	If the answers to both questions above are “no,” one may conclude that the intervention is not supported by meaningful evidence.

Note. Adapted from “Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide,” by U.S. Department of Education Institute of Education Sciences National Center for Education Evaluation and Regional Assistance, 2003.

The United States Department of Education (DOE) has also defined the two factors for EBP. The factors are: (a) the intervention may require strict adherence for the replication of the study within the school or the classroom and (b) the collection of outcome data may be needed to determine if the effects of the intervention in the study are being replicated in the classroom or school (US DOE, 2003). The DOE has determined that adherence to these two factors is reflective of a quality intervention.

Council for Exceptional Children and Evidence-based Practice

Odom et al. (2005) recognized that the key elements of EBP must be implemented in the profession of special education. Most taxonomies are designed to focus on RCT and acknowledge the importance of RCT. Then the targeted issue would be that RCT is not always the most suitable design for research in special education. Specifically, Odom et al. (2005) identified that certain aspects of research methodologies in special education

had to be clarified when not implementing RCT: (a) each study has to clearly match the research question and the methodology; (b) each of the four major research methods must be high quality; and (c) each use of the findings clearly relates to scientific evidence and best practices.

History of Taxonomies/Grading Systems

Many taxonomies have been developed with the primary focus on prevention of disease. The first taxonomy was developed and published in 1979 by the Canadian Task Force and known as The Periodic Health Examination. The two goals of this periodic health exam were: (a) to assess prevention practices of specific disease and (b) to increase health practices promotion. The Task Force wanted to evaluate two major factors: (a) the effectiveness of the intervention based upon the quality of the evidence and (b) the classification of the recommendation of the intervention. The levels of evidence would earn a ranking of I, II-1, II-2, or III. A classification of I is reflective of a RCT and was the highest and most favorable ranking, and a ranking of III was reflective of opinion or descriptive type studies and considered to be the least favorable. Each recommendation could earn a rank of A through E. A study earning a rank of A is identified as good evidence; and a study earning a rank of E is because of the exclusion of the treatment, and is the lowest recommendation (Canadian Task Force, 1979).

Users of Taxonomies

Primary users of taxonomies are in the profession of medicine and are responsible for assisting in determining the effectiveness of a treatment, effective techniques, and

publication of research results. Other professions currently using taxonomies include: dentists, diagnosticians, journal reviewers, nurses, therapists, prognosticators, and medical physicians (Baker, Potter, Young, & Madan, 2010; Kroke, Boeing, Rossnage, & Willich, 2004). Music therapists, speech therapists, and physical therapists are using taxonomies to better ascertain therapeutic practices that yield positive EBP outcomes. As professional professions have grown, so have the uses and purposes of taxonomies.

Purpose of Taxonomies

The intended use of taxonomies is to decrease bias and better identify EBP (Weiss, 2004). By comparing and contrasting existing research, a health care provider has the ability to determine if the outcome of the study is truly based on sound practices, establishes the benefits it claims, and/or if the results can only be replicated within a controlled population (Ebell et al., 2004). Another pertinent use of taxonomies is to identify areas of future research and exact areas of the paucity of research. Finally, the use of taxonomies also establishes a level of rigor for research, which to date has not been identified in the profession of Adapted Physical Activity.

Since the number of taxonomies available in various professional professions is increasing and taxonomies are being infused into published research, taxonomies are evolving. An example of the evolution of taxonomies is a shift in the focus from singular research designs to the incorporation of more than one research design. Another example of taxonomy evolution is the shift in focus from physician use only to patient centered taxonomy. The primary premise of the APAT is for use by professionals or those

studying the profession of Adapted Physical Activity. Therefore, the aim of this present research was to focus on the application of research and findings as opposed to patient centered taxonomies focused on treatment of a condition/disease.

Evolution of Taxonomies

As previously stated, the first taxonomy was developed by the Canadian Task Force (1979) and identified studies that were conducted with RCT. The Task Force did not identify any other types of research designs that could earn a strong rating for the effectiveness of an intervention and, therefore, RCT studies were identified as the only study to be of sound quality. Once effectiveness was established, the research was given a second rating based on its classification; however, if the study did not receive the highest recommendation for its effectiveness, it could not receive the highest rating for recommendation.

Since 1979, more than 100 taxonomies have been designed to rate the quality of a study and the recommendation of the treatment or practice have been published (Ebell et al., 2004). As taxonomies began to gain popularity, other shortcomings of these designs were identified. West et al. (2002) reported that many of these taxonomies were not comprehensive, and first designed only to evaluate one research design (i.e., randomized control trial, observational, meta-analysis, systematic review). A second shortcoming was that taxonomies were designed with the exclusivity of RCT as the only research design that could earn a high recommendation. A third shortcoming was that an

evaluator had to evaluate multiple taxonomies based on the specific research design which forced an evaluator to understand how to use multiple types of taxonomies.

Because of these shortcomings, experts began the development of comprehensive taxonomies in which practitioners can use one taxonomy to evaluate more than one research design. Another development in recent years has been the development of a number of taxonomies in multiple professions. For instance, within the medical community, nurses, dentists, and physicians all have different taxonomies designed to meet the unique needs of these professions and aligns with the standards of quality specific to unique needs of each profession. The final component is how a taxonomy can be used. The possibilities of its use are only now coming to the forefront as the quality of EBP is challenged.

Use of Taxonomies

Taxonomies have been implemented in various ways. Some taxonomies have been incorporated in: (a) dissertations, (b) meta-analyses, and (c) literature reviews of experimental research. The incorporation of a taxonomy has been presented in different formats. These formats may include in-depth table reviews, abbreviated table reviews of multiple studies, or included within the body of the text. The following are examples of various taxonomies from different professions and how a researcher may infuse a taxonomy into the publication of research.

Taxonomies within dissertations have been incorporated within the literature review to provide a critical, in-depth review of prior research. An example of how a taxonomy was

integrated into a dissertation is presented in Figure 1. In this example, a taxonomy was incorporated into a chart format which included findings of the study, statistical measures, type of study, and other information the researcher deemed pertinent. This information is used to provide an in-depth summary of the findings, provide an evidence-based score, and provide a recommendation of the findings. The evidence score in Figure 1 is noted by the circle and the recommendation is often located in the follow-up summary in the body of the literature. By incorporating a summary of findings, the reader, if he/she has knowledge of the taxonomy used to evaluate the research, can be informed of the quality of the supporting research.

Systematic Review of TPSR Literature: Hammond-Diedrich and Walsh

Auth /Lvel	Design/ Data collection	Sample	Outcome/ Intrvn/Context	Summary of Results
Hammond-Diedrich and Walsh (2006) 2	Case Study: Formal interviews Field notes Lesson Obs. Attendance records	Gender: 8 Males (N = 8) Demo: All Afr. Am. Ages (11-15years)	Outcome: Determine the potential impact of RM-based cross-age teaching program. Context: Participants were UYL that used prior RM experience to coach 40 4 th graders for eight sessions across five weeks who were involved in a summer sport program.	Qualitative: Emergent themes: (a) program expectations, (b) perceptions of the program, (c) progression of teaching skills, (d) familiarization of a university, (e) personal growth, & (g) responsibility development. Program provided advanced leadership opportunities, familiarization with a University developmental activities during summer discretionary time, continued contact w/ significant adults, & create relationships w/ peers from various programs across city.

Note. Auth = author; Lvel = level; Intrvn = intervention Demo = demographic; Afr.Am = African American; RM = Responsibility Model; UYL = Urban Youth League.

Figure 1. Example of dissertation use of a taxonomy, (adapted from Jackson, 2010).

Another use for taxonomies can be the implementation of a taxonomy in a meta-analysis or a formalized literature review study. In Figure 2, Guyatt, Oxman, and Schünemann et al. (2011) incorporated the use of a taxonomy when reviewing literature. This method of presentation is used to identify key information, which results in a

taxonomy score, and allows for a way of seeing all pertinent information for the analysis in one chart. When a taxonomy is implemented that is related to a meta-analysis, it provides the reader with the knowledge of the strength of the studies used to substantiate the research findings.

GRADE evidence profile: antibiotics for children with acute otitis media

Quality assessment						Summary of findings					
No of studies (Design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Number of patients		Relative risk (95% CI)	Absolute risk		Quality
						Placebo	Antibiotics		Control risk ^a	Risk difference (95% CI)	
Pain at 24h 5 (RCT)	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	241/605	223/624	RR 0.9 (0.78–1.04)	367/1,000	Not Significant	⊕⊕⊕⊕ High
Pain at 2–7 d 10 (RCT)	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	303/1,366	228/1,425	RR 0.72 (0.62–0.83)	257/1,000	72 fewer per 1,000 (44–98)	⊕⊕⊕⊕ High
Hearing, inferred from the surrogate outcome abnormal tympanometry—1 mo 4 (RCT)	No serious limitations	No serious inconsistency	Serious indirectness (because of indirectness of outcome)	No serious imprecision	Undetected	168/460	153/467	RR 0.89 (0.75–1.07)	350/1,000	Not Significant	⊕⊕⊕○ Moderate
Hearing, inferred from the surrogate outcome abnormal tympanometry—3 mo 3 (RCT)	No serious limitations	No serious inconsistency	Serious indirectness (because of indirectness of outcome)	No serious imprecision	Undetected	96/398	96/410	RR 0.97 (0.76–1.24)	234/1,000	Not Significant	⊕⊕⊕○ Moderate
Vomiting, diarrhea, or rash 5 (RCT)	No serious limitations	Serious inconsistency (because of inconsistency in absolute effects)	No serious indirectness	No serious imprecision	Undetected	83/711	110/690	RR 1.38 (1.09–1.76)	113/1,000	43 more per 1,000 (10–86)	⊕⊕⊕○ Moderate

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; RCT, randomized controlled trials; CI, confidence interval; RR, risk ratio.

^a The control rate is based on the median control group risk across studies.

Figure 2. Example of meta-analysis use of a taxonomy (Guyatt, Oxman, Schünemann et al., 2011).

A third example of displaying a recommended grade is provided in the designated circle of Figure 3. Similar to the presentation of the use within the dissertation, the recommended score may be located within the body of the findings or incorporated within a chart.

Work Group's Conclusion: Digitalis improves symptoms, exercise tolerance, and quality of life, but neither increases nor decreases mortality. Conclusion Grade: I						
Author/ Year	Design Type	Class	Quality +, -, \emptyset	Population Studied/ Sample Size	Primary Outcome Measure(s)/ Results (eg, <i>p</i> value, confidence interval, relative risk, odds ratio, likelihood ratio, number needed to treat)	Authors' Conclusions/ <i>Work Group's Comments (italicized)</i>
Captopril-Digoxin Research Group (1988) <i>continued</i>	RCT	A	\emptyset	<ul style="list-style-type: none"> - Concomitant therapy with inotropic agents, vasodilators, β-adrenergic blockers, calcium antagonists, immunosuppressive agents, or other investigational drugs was inotropic agents, vasodilators, β-adrenergic blockers, calcium antagonists, immunosuppressive agents, or other investigational drugs was prohibited 	<ul style="list-style-type: none"> - Similar trends seen in intention-to-treat analysis - Rate of discontinuation due to adverse drug reactions: 2.9% captopril, 4.2% digoxin, 0% placebo - More possible adverse drug effects attributed to captopril (44.2%) during blinded portion of study than to other treatments (30.2% digoxin, 24% placebo) (usually mild and transient dizziness and light-headedness) - 21 deaths (8 captopril, 7 digoxin, 6 placebo) 	NOTES: Trial was double-blind; did intention-to-treat analysis (as well as analysis while patients adhered to assigned therapy); most patients were NYHA functional class II.
Digitalis Investigation Group (1997)	RCT	A	\emptyset	<ul style="list-style-type: none"> - 6800 patients (302 clinical centers) with heart failure and left ventricular ejection fraction < 0.45 in normal sinus rhythm - 988 patients with heart failure and ejection fraction > 0.45 were enrolled in ancillary trial - May have been already receiving digoxin - Randomly assigned to digoxin or placebo (digoxin dose varied) - Other therapy used if patient had worsening symptoms of heart failure; if remained symptomatic, allowed open-label treatment with digoxin 	<ul style="list-style-type: none"> - Follow-up visits at 4 and 16 wks then every 4 mos (mean duration 37 mos, range 28-58 mos) - No significant differences between groups (baseline) - 1,181 deaths in digoxin group (34.8%), 1,194 in placebo group (35.1%) (RR = 0.99, 95% CI: 0.91-1.07) - 1,016 deaths from cardiovascular causes in digoxin group (29.9%), 1004 in placebo group (29.5%) (RR = 1.01; 95% CI: 0.93-1.10) - Trend toward lower risk of mortality attributable to worsening heart failure in digoxin group (<i>p</i> = 0.06) - 910 patients hospitalized for worsening heart failure in digoxin group and 1,180 in placebo group (RR = 0.72; 95% CI: 0.66-0.79) 	<ul style="list-style-type: none"> - In patients with left ventricular ejection fractions (\leq 0.45) digoxin had no effect on overall mortality when added to diuretics and ACE inhibitors; the risk of hospitalization was reduced and the combined outcome of death or hospitalization attributable to worsening heart failure was also reduced. In clinical practice, digoxin therapy is likely to decrease the frequency of hospitalization but not survival.

Conclusion Grading Worksheet: Congestive Heart Failure Guideline (continued)

Figure 3. Example of a taxonomy incorporating findings within a chart (adapted from Greer, Mosser, Logan, & Halaas, 2000).

In Figure 4, a third example of the current use of taxonomy is provided by Tsiros, Coates, Howe, Grimshaw, and Buckley (2011) who incorporated a taxonomy within the method section of an experimental research study. The purpose of including this information was to provide a critical review and synthesis of prior research and also provide an objective evaluation of each study.

Reference	Author, year, design, evidence level	Subjects*	Obesity definition, reference data	Outcomes	Absolute strength higher in OB?	Strength lower in OB?		Strength & weight status related?	Body size correction
						per kg mass	per kg FFM		
(24)	Maffiuletti 2008, CC, III-3	10 SOB, 10 non-OB males, 13–17 years	SOB BMI > 97%	BMI, Tanner, FFM (BI), Peak IK KE T, KE fatigue (Cybex)	Yes (IK KE T)	–	No difference	–	?Ratio
(18)	Almuzaini 2007, CS, IV	44 boys, 11–19 years	–	BMI, SF, Peak IK KE & KF T, IK KE endurance (Cybex)	–	–	–	Yes – BMI & strength ($r = 0.58-0.69$)	–
(17)	Duche 2002, CC, III-3	44 OB, 50 non-OB, –14 years	OB BMI > 97%, French	DXA (OB), SF (non-OB), BMI, CPP	Yes	Yes	No difference	Yes – CPP dependent on BF in OB	Ratio & allometric
(19)	Armstrong 2001, long., II	747, 10–11 years	–	SF, BMI, CPP, Tanner	–	–	–	Yes – mass explains CPP (0.88), SF & CPP ($r = -0.16$)	–
(20)	Grund 2000, CC, III-3	88, 4–11 years	OB > 97%, German	BMI, BI, KE + KF strength (CT)	Weakest group had higher BMI/FM	–	–	No	–
(15)	Ward 1997, CC, III-3	†54 OB, 96 non-OB, 5th grade girls	–	PA recall, BMI, SF, PWC170	Similar for W	Yes – W	Yes – W	–	Ratio
(14)	Blimkie 1990, CC, III-3	10 non-OB, 11 OB boys, 15–18 years	OB > 30% BF	SF, Tanner, IM & IK KE T (Cybex), EET, MUA, KE CSA	No difference	Yes	*No difference	–	Ratio
(13)	Blimkie 1989, CC, III-3	11 non-OB, 13 OB boys, 9–13 years	OB > 30% BF	SF, BMI, Tanner, R KE T (Cybex), EET, muscle CSA	No difference (KE)	Yes – KE	*No difference	–	Ratio

Tanner, the Tanner scale has been widely used to assess pubertal development (134), as body composition is known to vary with puberty. The Tanner scale depicts five stages of sexual maturation, represented by drawings of pubic hair, scrotum and testes and breasts and has been validated as a self-assessment measure (135).

*Unless otherwise stated subjects were of mixed gender.

†Only included African-American girls.

*Normalized using muscle cross-sectional area.

BF, body fat; BI, bioelectrical impedance; BMI, body mass index; CC, case-control observational; CPP, cycling peak power; CS, cross-sectional; CSA, cross-sectional area; CT, computer tensiometry; DXA, dual energy X-ray absorptiometry; EET, electrically evoked torque; F, flexor; FFM, fat-free mass; HGS, handgrip strength; IK, isokinetic; IM, isometric; KE, knee extensor; KF, knee flexion; long., longitudinal study; MUA, motor unit activation; non-OB, non-obese; OB, obese; PA, physical activity; PWC 170, physical work capacity cycle ergometry 170 test; SF, skin-folds; SOB, severely obese; T, torque; V, voluntary; W, power.

Figure 4. Example of use of a taxonomy in the method section of literature reviews within experimental research, (adapted from Tsiros et al., 2011).

The following is the final example of how taxonomies are currently implemented within the body of the text in a research article. Some researchers use this method within experimental research and others use this method within the literature review. Hahn (2009) provided an example of the embedded method of incorporating taxonomies into a research study: “The only available SORT level-1 trial of the population effectiveness of ICS in mild persistent asthma of recent onset is START, a large, multinational, randomized effectiveness trial in children and adults that did not exclude smokers” (p. 4). Other examples of an in-text taxonomy, such as Geerts et al. (2004), may appear following a statement regarding a finding: “**Recommendations: Trauma: 5.1.1.** We recommend that all trauma patients with at least one risk factor for VTE receive thromboprophylaxis, if possible (**Grade 1A**).” By providing a score in text with the quotation or citation of specific research, it provides the reader with the knowledge of the strength and quality of the research being used to substantiate the need for research.

Development and Use of Two Comprehensive Taxonomies

While the development of taxonomies are evolving, the framework developed by the Canadian Task Force (1979) concept still remains in many current taxonomies. The primary element of the Canadian Task Force taxonomy that is evident in other systems is dual ranking or classification (i.e., quality of research and recommendation). Each study is first evaluated based on the quality of the study; followed by a second ranking or classification, identified as the level of recommendation. In the medical profession, the recommendation refers back to the physician recommending a treatment to a patient that an intervention is effective, “yes,” “maybe effective,” “maybe with considerations,” or

“no” is not effective. In many current taxonomies if a study does not earn a top ranking in the quality of evidence it is very unlikely it can earn a high recommendation (Pilcher & Bedford, 2011).

The first known taxonomy to evaluate quality of study and level of recommendation was, Grading of Recommendations Assessment, Development, and Evaluation (GRADE [Guyatt, Oxman, Scünemann, Tugwell, & Knottnerus, 2011]). GRADE was the first known taxonomy that was developed for use by healthcare providers to guide in the decision making process of research-centered treatment. In 2002, the Agency for Healthcare Research and Quality began to develop and researched other comprehensive taxonomy. The comprehensive taxonomy was later named the Strength of Recommendation Taxonomy (SORT).

The following sections in this chapter will provide an overview of these two prominent taxonomies. A summary of how each taxonomy has been applied to research and the strengths and weakness of each will also be discussed.

Grading of Recommendations, Assessment, Development, and Evaluation

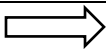
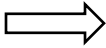
GRADE was developed to consolidate and provide a uniform taxonomy that could be used internationally to provide evaluative feedback on the quality of research. The aim was to eliminate the issue of who the consumer was within the medical profession so that there would be an understanding of what recommendations meant since everyone would be implementing the same taxonomy. The GRADE group identified a contingency of methodologists that included developers from other existing taxonomies as well as, major health organizations to form a committee to develop the GRADE taxonomy. The

GRADE group had 15 formalized meetings over 5 years and in 2004 the first GRADE taxonomy was published and distributed for use (American Thoracic Society, 2008). The overarching goal of the committee was to “. . . develop an optimal system of rating quality of evidence and determining strength of recommendations for clinical practice guidelines” (Guyatt, Oxman, Schünemann, Tugweel, & Knottnerus, 2011, p. 380).

Table 2 provides a summary of the use of the GRADE’s taxonomy with a recommendation of strong or weak and the quality of research is assessed based on high, moderate, low, and very low evidence (Guyatt, 2007). When a study is evaluated using GRADE, the evaluator is encouraged to consult the formalized evaluation document available on the GRADE website (<http://gradeworkinggroup.org/toolbox/index.htm>). A brief overview on how the evaluation process is conducted with GRADE consists of the reviewer first determining the quality of evidence. If a study is RCT, it starts on high quality and is then evaluated based on the negative and positive factors. All other research designs, including observational studies, are evaluated on low quality. Negative and positive factors are then applied.

Table 2

A Summary of Grading of Recommendations, Assessment, Development, and Evaluation's Approach to Rating Quality of Evidence Rating Symbol

Study design	Initial quality of a body of evidence	Lower if	Higher if	Quality of body of evidence
Randomized trials	High 	Risk of bias – 1 Serious – 2 Very serious Inconsistency – 1 Serious – 2 Very serious	Large effect + 1 Large + 2 Very large Dose response + 1 Evidence of a gradient All plausible residual confounding	High (four plus: ⊕⊕⊕⊕) or A Moderate (three plus: ⊕⊕⊕) or B
Observational studies	Low 	Indirectness – 1 Serious – 2 Very serious Imprecision – 1 Serious – 2 Very serious Publication bias – 1 Likely – 2 Very likely	+ 1 Would reduce a demonstrated effect + 1 Would suggest a spurious effect if no effect was observed	Low (two plus: ⊕⊕) or C Very low (one plus: ⊕) or D

Note. Adapted from “GRADE guidelines: 3. Rating the quality of evidence,” by Balshem, et al, 2011), *Journal of Clinical Epidemiology*, 64, 404.

Once the evidence of a study has been evaluated, it is assigned a symbol, as seen in

Table 3, to represent the overall quality of the evidence.

An example of use of current taxonomies applied to research in the profession of Adapted Physical Activity will follow the summary of the two most widely published

taxonomies as identified by the researcher. Two studies were randomly selected from the profession of Adapted Physical Activity and one from adapted physical education (APE). Each study was published in 2011 and evaluated using SORT and GRADE.

Table 3

Strength of Recommendation Rating Symbol

Strength of recommendation	Rating
Strong recommendation for using an intervention	↑ ↑ or 1
Weak recommendation for using an intervention	↑ ? or 2
Weak recommendation against using an intervention	↓ ? or 3
Strong recommendation against using an intervention	↓ ↓ or 4

Note. Adapted from “GRADE: going from evidence to recommendations,” by Guyatt, Oxman, Kunz, Falck-Ytter, Vist, Liberati, & Schünemann, (2008), *BMJ*, 336, 1051.

Strength of Recommendation Taxonomy

The Strength of Recommendation Taxonomy (SORT) was initially developed in 2002 and was “. . . intended to provide authors and readers of family medicine journals with a simple, user-friendly system for grading the strength of diagnostic treatment recommendations that appear in the articles in those journals” (Weiss, 2004, p. 141). The SORT began by comparing and contrasting existing taxonomies. Between SORT and GRADE taxonomies, SORT used one of the most comprehensive methodologies to develop a framework. Through an extensive literature search the SORT work group identified over 100 taxonomies (Ebell et al., 2004) and identified seven overlapping

developments. One distinguishing difference was that no one taxonomy incorporated or addressed all aspects of research design. The seven overlapping aspects were:

1. Instruments used to evaluate for specificity or general
2. Identified what type of instrument was used
3. Did the instrument define quality
4. What method was used to select items
5. What was the rigor of development
6. Established inter-rater reliability
7. Established clear instructions for use

Once evaluation of existing taxonomies were completed and the commonalities were identified, the committee then developed the framework for identifying common themes and the criteria for use within the medical community and leading medical organizations.

Table 4 contains an illustration of the uses of the SORT taxonomy.

Table 4

Assessing Quality of Evidence with SORT

Study quality	Diagnosis	Treatment/prevention/screening	Prognosis
Level 1: good-quality, patient-oriented evidence	Validated clinical decision rule SR/meta-analysis of high-quality studies High-quality diagnostic cohort study*	SR/meta-analysis or RCTs with consistent findings High-quality individual RCT [†] All-or-none study+	SR/meta-analysis of good-quality cohort studies Prospective cohort study with good follow up
Level 2: limited-quality patient-oriented evidenced	Unvalidated clinical decision rule SR/meta-analysis of lower quality studies or studies with inconsistent findings Lower quality diagnostic cohort study or diagnostic case- control study	SR/meta-analysis of lower quality clinical trials or of studies with inconsistent findings Lower quality clinical trial Cohort study Case-control study	SR/meta-analysis of lower quality cohort studies or with inconsistent results Retrospective cohort study or prospective cohort study with poor follow-up Case-control study Case series
Level 3: other evidence	Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented evidence (intermediate or physiologic outcomes only), or case series for studies of diagnosis, treatment, prevention, or screening		

Note. SR - systematic review, RCT - randomized controlled trial; *High-quality diagnostic cohort study: cohort design, adequate size, adequate spectrum of patients, blinding, and a consistent, well-defined reference standard; [†] High-quality RCT: allocation concealed, blinding if possible, intention-to-treat analysis, adequate statistical power, adequate follow-up (greater than 80 percent); + In an all-or-none study, the treatment causes a dramatic change in outcomes, such as antibiotics for meningitis or surgery for appendicitis, which precludes study in a controlled trial. Adapted from “JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews,” by Newman, Weyant, & Hujuel, (2007). *Journal of Evidenced Based Dental Practice*, 7,149.

Once the quality of the study is determined, it is further evaluated with another equation developed by SORT and then applied to the strength of recommendation grade as illustrated in Table 5. If a study is about origin (i.e., cause of a disease) it cannot receive a score any higher than level three, which translates into a final recommendation grade of C, the lowest level.

Table 5

Strength of Recommendation Grades

Strength of recommendation	Basis for recommendation
A	Consistent, good-quality patient-oriented evidence [*]
B	Inconsistent or limited-quality patient-oriented evidence [*]
C	Consensus, disease-oriented evidence [*] , usual practice, expert opinion, or case series for studies of diagnosis, treatment, prevention, or screening

Note. ^{*} Patient-oriented evidence measures outcomes that matter to patients: morbidity, mortality, symptom improvement, cost reduction, and quality of life. Disease-oriented evidence measure intermediate, physiologic, or surrogate end points that may or may not reflect improvement in patient outcomes (e.g., blood pressure, blood chemistry, physiologic function, pathologic findings). Adapted from “JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews,” by Newman, Weyant, & Hujoel, (2007). *Journal of Evidenced Based Dental Practice*, 7, 148.

As an example, following the summary of the two taxonomies, two studies were randomly selected from the profession of Adapted Physical Activity; one from APE and the other from disability sport. Each study was published in 2011 and evaluated to demonstrate how a study is evaluated with SORT.

Example of Two Prominent Taxonomies

Two research articles were randomly chosen when searching the Texas Woman's University library database, EBSCO. The search then was delaminated to databases within EBSCO were selected and publications from January, 2011 through December, 2011 were chosen. The first two research articles published in a scholarly journal were retained for application to GRADE and SORT. The quality of the study and the recommendation is presented in Table 6.

Table 6

Example of Quality of Study and Recommendation of SORT and GRADE Taxonomies

<u>Reference</u>	<u>Design</u>	<u>Method</u> <u>Instrument</u>	<u>Data Analysis</u>	<u>n</u>	<u>Participant</u> <u>M/F</u>	<u>Age</u>	<u>Diag.</u>	<u>Results</u>	<u>Taxonomies</u>
Zhang (2011)	Questionnaire	Questionnaire: 27th Annual Report to Congress Preparation and Job Demographics of Adapted Physical Educators in the United States	Estimation: Market-based analysis Prevalence-based projection model Difference between the two was then calculated	2 Data Sets	--	--	--	640 more APE teachers are need for current positions 20,087 APE teachers are needed to meet the need of students requiring APE Based on the difference there is a short fall of 66.81%	GRADE: ⊕↓↓ SORT: 3C FORM: DC
Breslin, C. & Rudisill, M. (2011)	3 testing conditions with randomization	Instrument: TGMD-2 Intervention: Visual supports	Within-subjects repeated- measures ANOVA	22	M/F	3.5- 10.92 years	ASD	Children with ASD gross motor quotient was more accurate with the use of visual supports	GRADE: ⊕⊕↓? SORT: 3C FORM: DC

Note. TGMD-2 = *Test of Gross Motor Deficiency-2*; n = number; M/F = Male and Female; Diag. = Diagnosis; ASD = Autism Spectrum Disorder; APE = Adapted Physical Education; ⊕ = Very Low; ⊕⊕ – Low; ↓? = Weak Recommendation Against an Intervention; ↓↓ = Strong Recommendation Against Using an Intervention; 3 = Other Evidence; C = Consensus, Disease-Oriented Evidence, Usual Practice, Expert Opinion, or Case Series for Studies of Diagnosis, Treatment, Prevention, or Screening; D = Studies, Or Level 1 To II Studies/Srs With A High Risk Of Bias; C = Body Of Evidence Provides Some Support For Recommendation(S) But Care Should Be Taken In Its Application

Based on these findings, neither of the selected studies can earn the highest quality of study based on the RCT requirement for each taxonomy. The study by Breslin and Rudisill (2011) could earn a stronger recommendation because it incorporated randomization into the study. However, because it lacked generalization, the recommendation would still be considered lower with some questions about application.

When evaluating these two studies against the Council for Exceptional Children's Practice Study Manual Indicators (2008) for EBP in special education, each of these articles would earn a coding of an insufficient evidence-base with potentially positive evidence base rating according to the score code. While RCT is considered the gold standard in medical research; in the profession of special education, many times the sample size is small, making it a challenge to obtain a large number of participants to conduct an RCT (Klingner & Boardman, 2011).

With the previous example there are drawbacks for using a taxonomy that is designed specifically for use within the medical community and there are limitations for crossing over into the profession of Adapted Physical Activity. The GRADE and SORT models also present strengths and limitations. The GRADE system primarily focuses heavily on RCT and meta-analyses (Baker, Potter, Young, & Madan, 2011). Within GRADE, recommendations can be cumbersome and difficult when outcome benefits are greater than the harm because it can still be graded as weak evidence. However, what makes this confusing is that a study can still earn a strong recommendation. In a study by García, Alvarado, and Gaxiola (2011), physicians who reported that using the GRADE system provided the most significant changes in clinical decisions. A study by Baker et al.,

(2010), however, it was reported that before there was a firm understanding and an ease of use of the GRADE system that users reported it to be cumbersome and confusing to understand.

In the SORT, RCT and systematic reviews are considered for high quality; all other types of research designs automatically receive a lower level grade (Faggion, 2010). The SORT system for evaluating evidence identified that if a study is disease-oriented, it will always receive a lower grade of A or B because of quality, quantity, and evidence (Faggion, 2010).

Council for Exceptional Children Quality Indicators

Because of the close affiliation of APE to special education, professionals in the profession of Adapted Physical Activity identify with organizations such as the CEC for research guidelines. During the past decade the CEC has established clear guidelines for determining the quality of specific research designs used in the special education literature. With the increase in testing, inclusion, and the use of best practices in special education, it has become imperative to those who implement new practices to be certain that research studies related to these practices are of high quality. Quality research should support recommendations for strategies that are sound, use the highest quality, and can be implemented even when individualizing instruction, which is a high priority to educators (Cook, Tankersley, & Landrum, 2009). The CEC has identified four basic research study designs with quality indicators. These designs are: correlational, qualitative, experimental/quasi-experimental, and single-subject designs (Odom et al., 2004).

Correlational

Thompson, Diamond, McWilliam, Snyder, and Snyder (2005) defined and outlined numerous indicators that lead to a strong correlational design. Correlational research is defined by Thompson et al. (2005) as “. . . quantitative, multi-subject designs in which participants have not been randomly assigned to treatment conditions” (p. 182) and variables are then compared against each other. The comparison of the variables is to determine if a relationship exists (Portney & Watkins, 2009). Odom et al. (2004) identified quality indicators for measurement, significance of the study, how to interpret outcomes, concerns of reliability, and the estimation of effect size. In Figure 5 the indicators are outlined.

<p>Measurement</p> <ul style="list-style-type: none"> • Score reliability coefficients are reported for all measured variables, based on induction from a prior study or test manual, with explicit and reasonable justifications as regards comparability of (a) sample compositions and (b) score dispersions. • Score reliability coefficients are reported for all measured variables, based on analysis of the data in hand in the particular study. • Evidence is inducted, with explicit rationale, from a prior study or test manual that suggests scores are valid for the inferences being made in the study. • Score validity is empirically evaluated based on data generated within the study. • The influences of score reliability and validity on study interpretations are explicitly considered in reasonable detail. <p>Practical and Clinical Significance</p> <ul style="list-style-type: none"> • One or more effect size statistics is reported for each study primary outcome, and the effect statistic used is clearly identified. • Authors interpret study effect sizes for selected practices by directly and explicitly comparing study effects with those reported in related prior studies. • Authors explicitly consider study design and effect size statistic limitations as part of effect interpretation. <p>Avoiding Some Common Macro-Analytic mistakes</p> <ul style="list-style-type: none"> • General linear model weights are interpreted as reflecting correlations of predictors with outcome variables only in the exceptional case that the weights indeed are correlation coefficients. • When noteworthy results are detected, and the origins of these effects are investigated, the interpretation includes examination of structure coefficients. • Interval data are not converted to nominal scale, unless such choices are justified on the extraordinary basis of distribution shapes and the consequences of the conversion are thoughtfully considered as part of result interpretation. • Univariate methods are not used in the presence of multiple outcome variables. • Univariate methods are not used post hoc to multivariate tests. • Persuasive evidence is explicitly presented that the assumptions of statistical methods are sufficiently well-met for results to be deemed credible. • Checks for Reliability Coefficients, Statistics, and Effect Sizes • Confidence intervals are reported for the reliability coefficients derived for study data. • Confidence intervals are reported for the sample statistics of primary interest in the study. • Confidence intervals are reported for study effect sizes. • Confidence intervals are interpreted by direct and explicit comparison with related CIs from prior studies.

Figure 5. Quality Indicators for Correlational Research. Adapted from “Quality Indicators for Research in Special Education and Guidelines for Evidence-Based Practices: Executive Summary,” by Odom et al., (2004). Council for Exceptional Children, Fall, 1-7.

Experimental/Quasi-Experimental

Horner et al. (2005) defined experimental and quasi-experimental designs as designs that “. . . allow researchers to determine whether implementation of a practice results in, or causes, a systematic change in specified outcomes within a defined population of students” (p. 150). Gersten, et al., further outlined quality indicators that not only would serve to develop research studies but can be used to evaluate and develop a quality proposal. Gersten et al., outlined quality indicators for participants, intervention, comparison condition, how outcomes are measured, and data analysis. Figure 6 outlines in detail quality indicators (Gersten et al., 2005).

<p>Quality Indicators for Describing Participants</p> <ul style="list-style-type: none"> • Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented? • Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions? • Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were compatible across conditions? <p>Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions</p> <ul style="list-style-type: none"> • Was the intervention clearly described and specified? • Was the fidelity of implementation described and assessed? • Was the nature of services provided in comparison conditions described? <p>Quality Indicators for Outcome Measures</p> <ul style="list-style-type: none"> • Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance? • Were outcomes for capturing the intervention's effect measure at the appropriate time? <p>Quality Indicators for Data Analysis</p> <ul style="list-style-type: none"> • Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study? • Did the research report include not only inferential statistics but also effect size calculation?
--

*Figure 6. Essential and Desirable Quality Indicators for Group Experimental and Quasi-Experimental Research Articles and Reports. Adapted from “Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education,” by Gersten, Fuchs, Coyne, Greenwood, & Innocenti (2005). *Council for Exceptional Children*, 71, 149-164.*

Qualitative

Brantlinger, Jimenez, Klingner, Pugach, and Richardson (2005) defined and outlined numerous indicators that lead to a strong qualitative research design. Qualitative research as defined by Brantlinger et al. has “. . . a systematic approach to understanding qualities, or the essential nature, of a phenomenon within a particular context” (p. 195). Because qualitative studies can be conducted with various methodological quality indicators, each is presented based on the method used. The most common qualitative research designs used in special education, according to Brantlinger et al., are: interview studies,

observational studies, and document analysis. Modification of these indicators from Brantlinger et al. are presented in Figure 7.

<p>Interview Studies</p> <ul style="list-style-type: none"> • Appropriate participants are selected • Interview questions are reasonable • Adequate mechanisms are used to record and transcribe interviews • Participants are represented sensitively and fairly in the report • Sound measures are used to ensure confidentiality <p>Observation Studies</p> <ul style="list-style-type: none"> • Appropriate setting(s) and/or people are selected for observation • Sufficient time is spent in the field • Researcher fits into the site • Research has minimal impact on setting • Field notes systematically collected • Sound measures are used to ensure confidentiality of participants and settings <p>Document Analysis</p> <ul style="list-style-type: none"> • Meaningful documents are found and their relevance is established • Documents are obtained and stored in a careful manner • Documents are sufficiently described and cited • Sound measures are used to ensure confidentiality of private documents <p>Data Analysis</p> <ul style="list-style-type: none"> • Results are sorted and coded in a systematic and meaningful way • Sufficient rationale is provided for what was included in the report • Documentation of methods used to establish trustworthiness and credibility are clear • Reflection about researchers' personal position/perspectives are provided • Conclusions are substantiated by sufficient quotations from participants, field notes of observations, and evidence of documentation inspection • Connections are made with related research
--

Figure 7. Quality Indicators Within Qualitative Research. Adapted from “Qualitative Studies in Special Education,” by Brantlinger et al, (2005). Council for Exceptional Children, 71, 195-207.

Single-Subject

Horner et al. (2005) defined and outlined numerous indicators that lead to a strong single-subject design. Single-subject research, as defined by Horner et al.), is “. . . experimental rather than correlational or descriptive, and its purpose is to document causal, or functional, relationships between independent and dependent variables”

(p. 166). Key elements of the quality indicators of this research design (Horner et al., 2005) are presented in Figure 8. These indicators address the major components of most single-subject designs: participant descriptions, setting descriptions, identification of the dependent and independent variables, baseline, control and validity (i.e., internal, external, and social). While exceptions do exist for the omission of some components due to the specific nature of single-subject research, Horner et al. emphasized that a quality single-subject design study will exhibit many indicators (see Figure 8).

Description of Participants and Settings
<ul style="list-style-type: none"> • Participants are described in detail • Participant selection is described in detail • Clear picture of physical setting is described
Dependent Variable (DV)
<ul style="list-style-type: none"> • Described with operational precision • Each variable is measure with quantifiable index • DV is replicable • DV are measured repeatedly over time • Reliability and inter-observer agreement for each DV are measure and meet standards
Independent Variable (IV)
<ul style="list-style-type: none"> • IV is replicable • IV is manipulated and under the control of experimenter • Fidelity of implementation for the IV is highly desirable
Baseline
<ul style="list-style-type: none"> • The study will include a baseline phase the provides repeated measurement of a DV and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the IV did not occur. • Baseline conditions are replicable
Experimental Control/Internal Validity
<ul style="list-style-type: none"> • At least three demonstrations of experimental effect at three different points in time • Controls for threats to internal validity • Results demonstrate experimental control
External Validity
<ul style="list-style-type: none"> • Experimental effects are replicated across participants, settings, or materials to establish external validity.
Social Validity
<ul style="list-style-type: none"> • DV is socially important • Magnitude of change of DV is socially important • Implementation of IV is practical and cost effective • Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social context.

Figure 8. Quality Indicators Within Single-Subject Research. Adapted from “Qualitative Studies in Special Education,” by Horner, Carr, Halle, McGee, Odom, & Wolery, (2005). Council for Exceptional Children, 71, 165-179.

Following other professional professions, and recognizing that EBP is crucial to the identification of sound research practices, the development and implementation of an taxonomy that is reflective of the profession of Adapted Physical Activity is paramount. Incorporating the quality of research designs by the American Psychological Association, the CEC, and adopting practices that are reflective of EBP according to the DOE, will yield an effective instrument. This instrument in turn should provide Adapted Physical

Activity researchers with a taxonomy that will effectively evaluate the quality of research to be published and evaluate published research.

CHAPTER III

METHOD

Because an Adapted Physical Activity Taxonomy (APAT) does not exist, taxonomies previously published from other professions were used to guide the investigator in the development of the APAT. The purpose of this investigation was fourfold. The first purpose of this investigation was to determine through a systematic search of literature what research taxonomies exist in the English language, are used in any profession, and are embedded within published literature. A second purpose was to identify categories and criteria present within the published research taxonomies.

Based on the findings of the literature search, the third purpose was to identify categories within the research articles and what criteria are needed in the APAT. The fourth purpose was to determine the inter-rater reliability (IRR) of the categories and criteria included in the APAT that were developed.

This chapter provides a description of the participants, instrumentation, procedures, research design, and data analyses for the development of an APAT. A meta-framework for instrument development, fidelity, and construct validity were used which consists of the first five phases (see Figure 9) and were grouped into the three sequential studies (Onwuegbuzie, Bustamante, & Nelson, 2010).

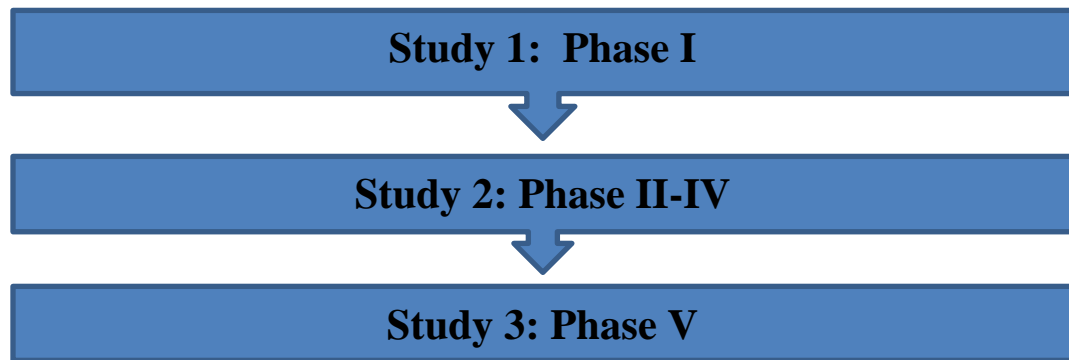


Figure 9. Study and Phases.

General Research Design

A modified meta-framework is a strong, logical study design according to Johnson, Onwuegbuzie, and Turner (2007) that: (a) exhibits both aspects of data collection and answer questions being tested; and (b) exhibits development that goes beyond the setting of the study (Johnson et al., 2007).

Onwuegbuzie et al. (2010) outlined a 10-phase process used in the meta-framework design (see Figure 10). Five of the 10 phases in the present investigation were grouped into 3 sequential studies that incorporate the first 5 phases of the modified meta-framework:

- Study 1 consisted of Phase I - Literature analysis;
- Study 2 consisted of Phases II - Delphi Survey; Phase III - APAT development; Phase IV - APAT pilot study; and
- Study 3 consisted of Phase 5 - Inter-rater reliability.

For the purpose of this dissertation, each of the 3 studies (i.e., not including Phases VI-X) are explained in detail and included the method and statistical analysis for each study.

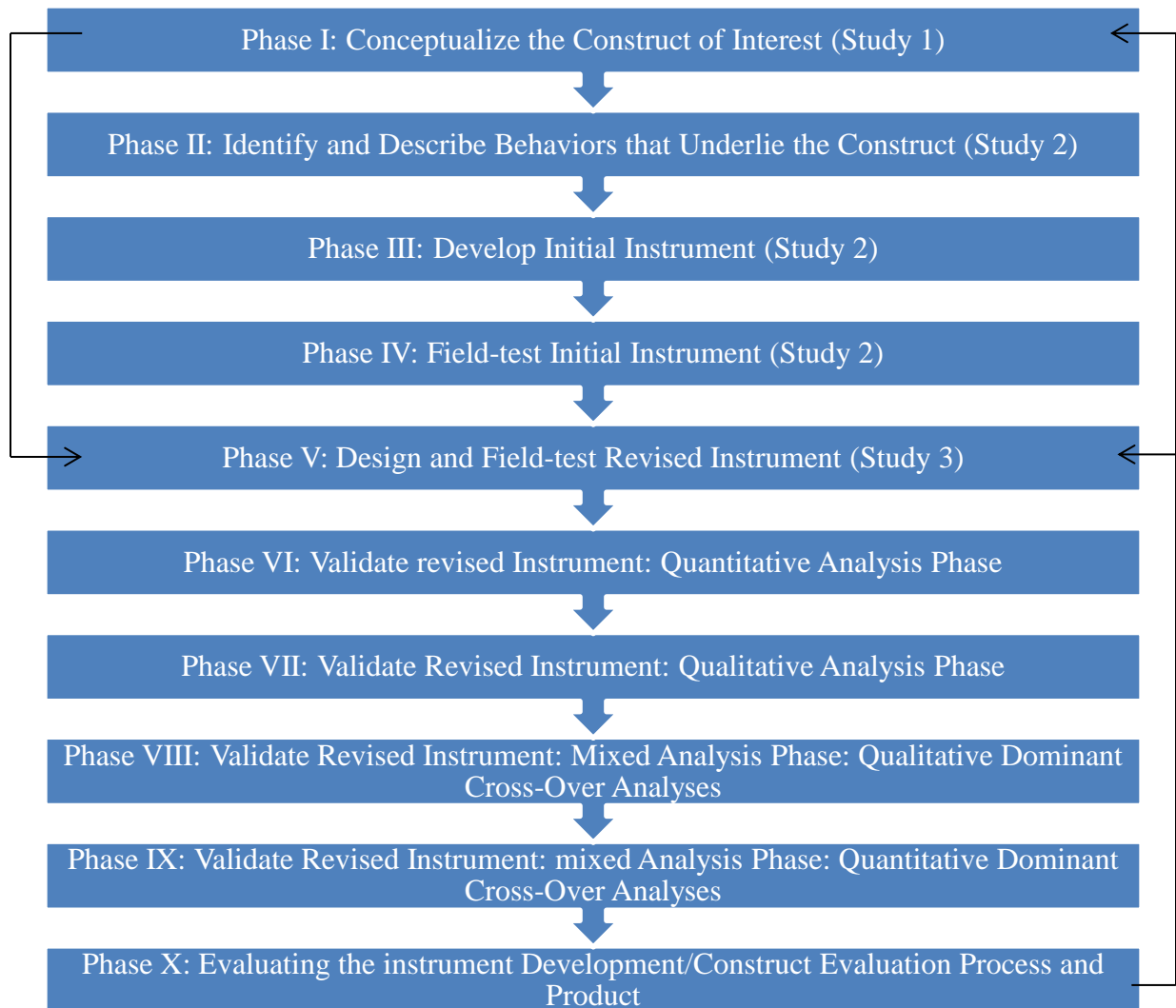


Figure 10. Instrument development and construct validation process. Adapted from “Mixed Research as a tool for developing quantitative instruments,” by A. J. Onwuegbuzie, R. M. Bustamante, & J. A. Nelson, (2010). *Journal of Mixed Methods Research*, 4, 62. (used with permission).

Study 1: Phase I – Literature Analysis

Study 1 consisted of an extensive examination of literature to locate research taxonomies and does not include participants. This study was conducted to identify research taxonomies embedded in published research. A systematic search was used to

collect existing journal articles in the areas of pediatrics, dentistry, nursing, nutrition, speech and language, healthcare, and music therapy. The criteria for the literature search were: (a) embedded research taxonomy within a research article, meta-analysis, or literature review; (b) peer reviewed articles located in a journal; (c) available in English; (d) published between 1979 - 2012; (e) the original sources, located in a research or scholarly article; and (f) all available in full text from the Texas Woman's University library. The year, 1979, was selected because the Periodic Health Examination (Canadian Task Force) is the first known research taxonomy to be published in the English language.

For this study the researcher:

1. Identified the number of current research taxonomies.
2. Determined if any other professions were identified through the literature review.
3. Determined which research designs the research taxonomy were intended to evaluate (e.g., multiple research designs, single-subject designs).

The retrieved research articles were used in this investigation through the following procedures. Of the 165 plus library databases at Texas Woman's University (TWU), 10 were recommended by a TWU Library Reference Specialist for this investigation. Of the 10 recommended databases, the library specialist then identified the following three databases to search based on specific parameters that are not available in the other seven databases: (a) *Medline* with Medical Subject Headings (MeSH) parameters, (b) *Current Index to Nursing and Allied Health Literature* (CINAHL) with CINAHL headings, and (c) *Psychological Information Database* (PsycINFO) with thesaurus options.

A search was conducted in the remaining seven databases with the terms identified from a search in CINAHL, Medline, and PsycINFO. The remaining databases to search were *Cochrane*, *Educational Resources Information Clearinghouse (ERIC)*, *Journals of Ovid*, *Nursing*, *Occupational Therapy Search*, *ProQuest*, *Scopus*, and *Sports Discus*.

Based on the findings of these searches, the researcher identified commonalities that included the following categories: (a) types of research designs, (b) standards used to develop the research taxonomy, (c) research taxonomy evaluative criteria in categories, and (d) research taxonomies used to evaluate a research design. Each of the categories were summed for a serial count in order to identify the most common theme. The literature identified in Study 1 was reviewed and the information gathered in the categories, as stated earlier, was used in Study 2 to develop the Delphi Survey.

Study 2: Phase II - IV

Phase II: Delphi Survey Technique

The purpose of Study 2 was to determine the key criteria (i.e., key statements) that are needed in an Adapted Physical Activity taxonomy. Twelve purposefully selected professionals were initially recruited for Study 2 based on the following criteria: (a) were professional working in the area of Adapted Physical Activity, (b) were members of the National Consortium for Physical Education for Individuals with Disabilities, (c) were an associate professor or professor (retired or active), and (d) previously consented to volunteer in the study (Belanger & Colantonio, 2011; Okoli & Pawloski, 2004; Thangaratinam & Redman, 2005).

This study involved the use of the Delphi Survey method which is “. . . a multistage process designed to combine opinion into group consensus” (Hasson, Keeney, & McKenna, 2000, p. 1010). This multistage process incorporated the findings of Study 1 to develop the first round of Delphi questions (see Appendix B). Questions were developed to interpret the following statements in the first round of Delphi questions:

1. Rank the type of research designs in order of importance;
2. Identify a specific set of standards to be used to guide the development of an Adapted Physical Activity taxonomy;
3. Analyze the research article (i.e., general, design, outcome) for category development;
4. Identify the criteria that should be included within each category; and
5. Providing an opportunity for participants to provide open ended feedback with the survey.

After the first round of questions were developed, three professors of Adapted Physical Activity and four doctoral students in Adapted Physical Activity completed the survey checking for clarity of the questions and length of time to complete the survey. The survey was then administered.

In order to minimize research bias and to achieve a high level of trustworthiness of the data, the researcher used a method of leaving an “audit trail” (Akkerman, Admiraal, Brekelmans, & Oost, 2008). Data were prepared by organizing responses of participants on a spreadsheet. Three graduate students of Adapted Physical Activity, who had completed no less than 3 hours of a statistics or research methods course, audited each

transcription of the first survey for accuracy. The graduate students were provided the original surveys returned and confirmed the transcription of the data onto the spreadsheet. Following the audit the review for IRR and coding of data from participants responses were conducted. Coding of data was completed by two individuals: (a) a member of the dissertation committee and (b) a person with a PhD in Adapted Physical Activity, who did not participate in the survey. Audit procedures described by Akkerman et al. (2008) were implemented and are as follows:

1. Orient the auditors to the procedures of the study
2. Orient the auditors to the study
3. Orient the auditors to their role in coding data reliability

Each auditor was provided a copy of the raw data. The auditors then read each response and highlighted the key terms and/or phrases from the surveys (Akkerman, Admiraal, Brekelmans, & Oost, 2008; Rodgers & Cowles, 1993). Based on the results of the identified key terms and/or phrases, two themes were identified (Akkerman, Admiraal, Brekelmans, & Oost; Rodgers & Cowles). Each of these terms and/or phrases were then coded, the IRR obtained, and considered acceptable if greater than .75 (Portney & Watkins, 2009). Each of the coded open-ended questions were then turned into Likert-type questions for round two of the Delphi Survey (Linstone & Turoff, 2002).

Once consensus was obtained by no less than 7 of the 12 participants (as cited in Linstone, 1978) in the first round of questions, subsequent questions were developed to address and clarify areas of concerns. Once consensus was obtained, participant's

responses from the Delphi Survey were used to develop categories and criteria for the initial formatting of an Adapted Physical Activity Taxonomy.

Phase III: Adapted Physical Activity Taxonomy Development

This phase was based on the results obtained in Phase I and II of Study 2. The format for the APAT was organized and format based on the SORT. The format was based on the taxonomy with the highest number of articles with a taxonomy imbedded, this was the SORT. The organization of content was developed based on the agreed statements from the participants collected in Phase II of Study 2. The Journal Article Reporting Standards format from the American Psychological Association (2010) was used as the format for the categories. The initial APAT was designed based on the format of the most widely published and embedded research taxonomy identified in Study 1, SORT. Once the initial APAT was developed with categories and criteria, it was reviewed with an opportunity for open-ended written feedback from the participants in Study 2. Feedback provided from each participant was then evaluated and edits were implemented if it met the previously identified criteria that align with the Council for Exceptional Children (CEC) and the American Psychological Association for quality researcher; or was considered to be general edits (i.e., spelling, grammar, or flow).

Phase IV: Pilot Study

Once the categories and format were determined in Phase III, the criteria were placed into the categories. Based on Onwuegbuzie et al. (2010), items should be assessed for clarity, relevancy, and competency. To accomplish this, the initial instrument was tested using 10 doctoral students in Adapted Physical Activity who were asked to evaluate a

purposefully selected research article. The purposefully selected article met the following criteria: (a) it is available in full text; (b) written in English; (c) falls under the search perimeter for Adapted Physical Activity; and (d) the study was conducted using one of four research designs (i.e., correlational, experimental/quasi-experimental, qualitative, and single-subject). For the correlational design, the article by Tseng, Chen, Shieh, Lu, and Huang (2011) was selected. For experimental/quasi-experimental the article by Willoughby, Dodd, Shields, and Foley (2010) was selected. For qualitative an article by Wilski, Nadolska, Dowling, McConkey, and Hassan (2012) was selected. Then for single-subject design a study by Brien and Sveistrup (2011) was selected. Participants were asked to read the instructions and use the APAT to identify the Strength of Quality (SOQ) and the Level of Recommendation (LOR) of the study. Participants were also given the opportunity to provide written feedback regarding the clarity, ease of use, and functionality of the instrument.

Study 3: Phase V

Inter-Rater Reliability

The fourth purpose of this investigation was to identify the reliability of the APAT which includes the SOQ and LOR, developed in Study 2. In this investigation, websites in English were searched to identify universities in the United States and Canada and to recruit no less than 60 participants from with Adapted Physical Activity programs or professors with a background in Adapted Physical Activity. Participants were also recruited from the Certified Adapted Physical Educators database, as well as, the National Consortium for Physical Education for Individuals with Disabilities.

Participants in Study 3 were tenured, associate professors (active or retired), work in the United States or Canada, and have completed graduate course work in Adapted Physical Activity (and were not the same as those used in Study 2). The number of participants needed were determined by conducting an a priori power analysis using G*power (Faul, Erdfelder, Buchner, & Lang, 2009), specifying an effect size of .3 and β set at .8.

Study 3 involved inter-rater reliability analysis of the APAT. Each participant was invited by e-mail to consent and participate in the study through Psychdata. Consenting participants were directed to Psychdata to read a preselected research article based on one of four research designs (correlational, experimental/quasi experimental, qualitative, and single-subject). Participants were asked to use the APAT to score an article in Psychdata. Each research design in Psychdata was set up as an individual survey and designed using a method known as skip logic. Using this method, participants were asked to provide a score for SOQ and for LOR. Scores were then compared to the findings of the group for each research design independently. At the conclusion of the initial evaluation participants were given the option to provide feedback and complete another evaluation based on a different research design with the possibility of completing all four research designs.

CHAPTER IV

RESULTS

This investigation was designed to answer three research questions: (a) What are there essential criteria that need to be included in an Adapted Physical Activity taxonomy?; (b) What categories and criteria can be determined through the use of a validation process?; and (c) What level of inter-rater reliability of an Adapted Physical Activity taxonomy can be established?

In order to answer these three questions, four purposes of this investigation have been developed. Within this investigation, there are four studies and each study involves one or more phases of the Instrument Development and Construct Validation (IDCV) process presented in Chapter 3. The results for each study and corresponding phase(s) will be presented. The subheadings of this chapter are presented in Figure 11.

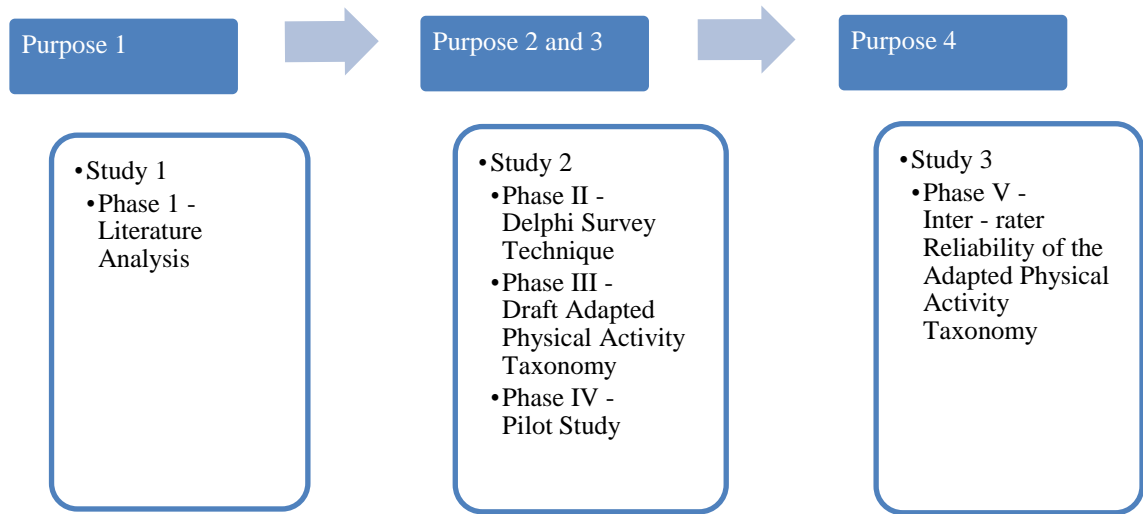


Figure 11. Visual representation of purposes, three studies, and five phases of the investigation.

Study 1: Literature Analysis - Phase I

The first purpose of this investigation was to determine through a systematic search of literature what research taxonomies exist in the English language, are used in any profession, and are embedded within published literature. Study I comprised one phase. In Phase I of this study there was an intensive analysis of literature to: (a) identify the number of taxonomies, (b) identify the professions using taxonomies, (c) identify research designs implemented through the use of taxonomies, and (d) identify taxonomies that were imbedded in published literature between 1979 and 2012. After the completion of Phase I, the investigator was able to determine and develop the Delphi Survey questions to be used in Study II. The presentation of these findings will be presented in a sequential order base on the findings of the analysis of the literature.

Literature Analysis

To locate research taxonomies, the researcher analyzed articles that included abstracts, to identify taxonomies that were imbed in the literature. The investigator identified a total of 2,607 articles meeting the search criteria. The two databases that produced the largest number of articles were *Medline* with Medical Subject Headings (MESH) and *Current Index to Nursing and Allied Health Literature* (CINAHL) with headings. In Table 7 a complete list of all databases searched and number of articles retrieved is presented. A complete listing of databases searched, search terms used, and number of articles are available on request. All articles were then saved in an excel spreadsheet and duplicate articles were removed.

Table 7

Database and Numbers of Located Articles

Database searched	Number of articles located
CINAHL with Headings	636
Cochrane	95
ERIC	15
Journals of Ovid	36
Medline with MESH	1,233
Occupational Therapy Search	55
ProQuest Nursing & Allied Health Source	166
PsycINFO with Thesaurus	72
Scopus	99
Sports Discus	200
Total	2,607

There were 2,607 articles identified on a search based on specific word(s)/term(s). Only 67 articles met the following criteria: (a) were embedded research taxonomy within a research article, meta-analysis, or literature review available in full text including an abstract; (b) were peer reviewed articles; (c) were available in English; (d) were published between 1979 to 2012; and (e) were the original sources located in a scholarly journal. The 67 remaining articles were then analyzed through three leading qualifiers listed below to identify the taxonomies that were imbedded in articles meeting the criteria above:

1. Identify the number of different research taxonomies that have been included within the 67 articles in published research - Purpose 1:

From the search of the 67 articles, the top three ranking taxonomies (highest number of imbedded articles) and the number of articles imbedding a taxonomy within the article were identified: (a) Strength of Recommendation Taxonomy (SORT); (b) Grading of Recommendation Assessment, Development and Evaluation (GRADE), and (c) the National Institute for Clinical Excellence (NICE). A complete listing of all of the other taxonomies identified and discussed within the 67 articles and the number of articles imbedding the specific taxonomy are available on request.

2. Determine if any other professions than those previously identified have a research taxonomy - Purpose 1:

In a previous investigation, the professions that use taxonomies were: dentistry, healthcare, music therapy, nursing, nutrition, pediatrics, and speech and language (Belanger & Colantonio, 2011). By investigating the taxonomies identified in Qualifier 1

these professions were confirmed. The investigator also identified the number of publications (i.e., both refereed and non-refereed) in which taxonomies were infused within published research. In rank order, mentioned at least three times in the literature, five other professions were identified: general medical (n = 31), psychology (n = 7), pediatrics (n = 5), health care (n = 4), orthopedics (n = 4), and others (e.g., physical and music therapy; n = 16) for a total of 67. A comprehensive list of the professions identified imbedding research taxonomies and number of articles located for each is available on request.

3. Determine the types of research designs and the categories used within the designs in the identified 25 taxonomies (e.g., multiple research designs, single-subject designs, qualitative) - Purpose 1:

Of the 67 articles identified in qualifier 1, there were 25 different taxonomies identified that were embedded in the 67 articles. The researcher then located the information (as stated above) about each of the specific taxonomies in qualifier 1. For example, in order to determine how SORT was developed, the researcher searched the same databases previously suggested by the research librarian of Texas Woman's University for information regarding the development of the taxonomy. The researcher also searched the web browsers Bing, Dogpile, Google, and Google Scholar to locate possible information on the development for SORT and was conducted for all taxonomies that were identified in Qualifier 1. Based on the results of this search, of the 25 taxonomies identified, 12 of the taxonomies addressed multiple research designs, while in

13 taxonomies no specific research design(s) were identified as described above. A summary of the research designs are located in Table 8 below.

Table 8

Number of Research Designs Evaluated

Count	Research Design
41	Systematic Review
20	Literature Review
1	1 Sys Rev/Meta
1	1 Survey
1	1 Meta-analysis
1	1 Qualitative
1	1 Reflective Critique (Literature Review)

When investigating the development of categories within the 25 taxonomies, 11 taxonomies were identified using clear category headings, 8 did not specify or could not be located, 3 were developed by standards from specific profession of emphasis (i.e., Obstetrics and Gynecology from specific standards to the profession), and 3 were based on a survey process. The survey process identified was Delphi Survey, questionnaire, and the third did not specify. By answering the three qualifiers listed above, the results of the literature analysis were used to develop the first round of Delphi Survey questions to be used in Study 2. Additional information on the 67 articles related to not only the research design type, but how the present taxonomy were developed, and taxonomy categories are available on request.

Study 2: Phases II-IV

The second purpose of this investigation was to identify categories and criteria present within the published research taxonomies based on findings in Study 1. The third purpose of this investigation was to identify categories within the research articles and what criteria are needed in the Adapted Physical Activity Taxonomy (APAT) based on the findings in Study 1. Study 2 comprised of three Phases (i.e., Delphi Survey, APAT Development, and APAT Pilot Study). In Study 2 of Phase II, purposefully selected participants were asked to participate in the Delphi Survey technique. This survey method was used to bring participants to a consensus on what criteria (i.e., Council for Exceptional Children, SORT) should be used in the APAT, what research designs the APAT would evaluate, and what categories and criteria should be included groupings in the APAT. In Phase III the same participants used in Phase II were asked to provide feedback on the draft version of the APAT developed during the Delphi Survey method in Study 1. The fourth and final phase of Study 2 was to Pilot Test the APAT draft using six Adapted Physical Activity doctoral students.

Delphi Survey Technique Phase II

Phase II, in Study 2, was designed to determine the categories and criteria needed in the APAT. There were 12 purposefully selected professionals recruited for Study 2 based on the following criteria: (a) professionals working in the area of Adapted Physical Activity, (b) members of the National Consortium for Physical Education for Individuals with Disabilities, (c) associate or full professor (retired or active), and (d) previously consented to volunteer in this study (Belanger & Colantonio, 2011). This study involved

the use of the Delphi Survey method and involved 3 rounds. Subsequent rounds would have been administered if 75% agreement among the participants was not met (Hasson, Keeney, & McKenna, 2000). The findings for each survey round results were as follows:

Round 1. The Delphi Survey 1 consisted of 11 questions (see Appendix B). This survey was sent to the 12 participants; 8 were returned completed. In an attempt to minimize research bias, the researcher implemented a method of leaving an “audit trail” in which 3 graduate students in the profession of Adapted Physical Activity reviewed raw data of Survey 1 to ensure the accuracy of data transfer to the Excel spreadsheet during Round 1 of the Delphi Survey. The first survey question for all rounds of the surveys asked participants to state their name; the response to this question was omitted in order to maintain confidentiality (see Table 8). Questions 2 through 5 were Likert-type demographic questions. The results of the 8 returned surveys are presented in Table 9.

Table 9

Participant Demographics

Likert-type questions	Results
<u>Question 2</u>	6 participants = 15+ years
Total number of years as Associate	1 participant = 3-6 years
Professor or higher	1 participant = 0-3 years
<u>Question 3</u>	5 = full professors
Current academic rank	3 = associate professors
<u>Question 4</u>	4 = 16+ data-based publications
Number of data-based refereed publications	1 = 9-15 data-based publications
as first author	1 = 6-9 data-based publications
	1 = 3-6 data-based publications
	1 = 03 data-based publications
<u>Question 5</u>	2 = 16+ non-data publications
Non-data refereed publication as first author	4 = 10-15 non-data publications
	1 = 7-9 non-data publications
	1 = 4-6 non-data publications

Question 6 was a Likert-type question where participants ranked which criteria should be used to develop the APAT. The results of the rankings are as follows from the highest to the lowest: (a) the American Psychological Association; (b) the Council for Exceptional Children Research Indicators; (c) the Strength of Recommendation Taxonomy; (d) the Grading of Recommendation, Assessment, Development, and Evaluation; and (e) the PhD Competency Guidelines. There were no other standards suggested by the participants.

Question 7 was also a Likert-type Question where participants ranked which research designs used in Adapted Physical Activity were the *most important* to them as a researcher. A ranking of 1 indicated most important to 7 being least important. The results of the ranking are: (a) mixed method and quantitative; (b) quasi-experimental; (c) single-subject; (d) correlational; and (e) qualitative. During this process participants had a high level of interaction with the investigator that included e-mails and telephone calls to clarify information (i.e., quantitative design grouped into experimental/quasi-experimental).

In addition, Question 8 was a Likert-type question where participants ranked which research design they *used the most* as a researcher. A ranking of 1 indicated most important to 7 being least important. The results of the ranking are as follows from highest to lowest: (a) mixed method and quantitative design; (b) quasi-experimental; (c) correlational; and (d) qualitative. Single-subject was not a research design chosen by any of the participants.

Question 9 was an open-ended question asking participants to identify and briefly describe what information they would expect to find within the Introduction, Method, Results, and Discussion portion of a manuscript. Participants were also asked if there was any “other” information that should be included that were not identified into these categories. For the remainder of the study this category is referred to as “Other.” All the unedited responses from the participants are available on request.

Question 11 was also an open-ended question. The participants were asked what they would consider before developing a study to help address the level of evidence (e.g., a

study design, statistical analysis, participant selection). The unedited responses to these questions are available on request.

The responses to Questions 9 and 11 were entered into Nvivo 8 (2012) to perform a word analysis and identify high frequency words. The investigator and two individuals with a PhD in Adapted Physical Activity were auditors. The roles of the auditors are described on page 62. Each auditor was provided a copy of the raw data. The auditors then read each response and highlighted the top two themes, key terms and/or phrases from the surveys based on a process of selective coding provided through Nvivo 8. Selective coding allowed for the researcher to identify clear themes and/or key terms in the participant's answers in Study 2, Delphi Survey Round 1 (Pitney & Parker, 2009).

The Pearson correlation coefficients for the participant responses were obtained from the coded data by the researcher and the two auditors. The correlation coefficient was considered acceptable if it was between .75 to 1.0 and -.75 to -1.0 (see Table 10). Because of the discrepancy in participant responses, the "Results Section" and "Discussion Section" did not meet the acceptable level for the correlation coefficient. Each of the coded open-ended questions (each part of Question 9 and 11) was then turned into Likert-type Question for Round Two of the Delphi Survey (Linstone & Turoff, 2002).

Table 10

Nvivo Analysis of Correlation Coefficient for Participant Responses

Question	Section	Pearson correlation coefficient
9	Introduction	0.925
9	Method	0.962
9	Results	0.196
9	Discussion	-0.711
9	Other	0.999
11	Level of evidence	0.911

Parts of Question 9 and Question 11 meet the criteria with a very high correlation coefficients. The “Results” and “Discussion” parts of Question 9 by the participants indicated future area of concern. Because this portion of Question 9 did not have a high correlation coefficient, the content of this question needed to be addressed in Survey 2.

Participants were asked in Question 10 if they “strongly agree” to “strongly disagree” to the following statement: “strength of a recommendation indicates the extent to which we [researchers] can be confident that the recommendation will do more good than harm...” The results to this question were: (a) strongly agreed (n = 3), (b) somewhat agreed (n = 3), and (c) did not answer (n = 2).

Round 2. After analysis in Round 1 the responses to the open-ended questions were then converted to Likert-type questions. Based on the findings from Delphi Survey 1, Likert-type questions were revised for Delphi Survey 2 (see Appendix C). Round 2 was needed because the first round consisted mostly of open-ended questions. A total of 8

participants returned the first survey therefore, 8 surveys were administered in Delphi Survey 2. In Delphi Survey 2, 1 participant dropped out resulting in the return of 7 completed surveys for Round 2.

Consensus was obtained with agreement of 70% or higher on all parts of Questions 2 through 6. Question 2 was a follow-up question from Survey 1 regarding clarification for three research designs having a disparity in Round 1 (i.e., correlational, single-subject, and qualitative). Question 3 was a follow-up question from Round 1 regarding what information should be included within the “Introductory Section” of research articles. Question 5 was a follow-up question from Round 1 regarding what information should be included within the “Results Section” of a research article. Question 6 was a follow-up question from Round 1 regarding what information should be included in the “Discussion Section” of a research article. Because a consensus was met, no further questioning was needed.

Because a 70% agreement was not met for Questions 3 or 7 a Round 3 was needed. Question 3 related to the “overview of the study design” with a 50% agreement and “solutions to challenges” with 50% agreement. Question 7 related to information that should be included within the “Other Section” of the APAT there was a 67% agreement. The three items not meeting the 70% agreement were further addressed in Delphi Survey 3.

Round 3. The third round of questions was sent to the 7 remaining participants to address the statements of Question 3 and 7 that did not meet the .70% agreement in Survey 2. Using Likert-type questions, participants responded to what statements they

agreed or disagreed. Question 2, in Delphi Survey 3, was a follow-up to the “Introductory Section” and participants were asked about the inclusion or exclusion of the “overview of the study design” and “solutions to the challenges.” Participants determined that the “overview of the study design” should be excluded while “solutions to challenges” should be included.

Question 3 in Survey 3 was a follow up to the “Other” section and participants were asked about the inclusion or exclusion of “appendices.” Participants determined that the “appendices” should be included. Consensus of .70 was met and this concluded the Delphi Survey process. Participant’s responses from the Delphi Survey were then used to develop the draft version of the APAT.

Adapted Physical Activity Taxonomy Development Phase III

This Phase was based on the results obtained in Study 1, Phase I and Study 2, Phase II. The APAT draft was designed based on the SORT format which is the most widely published and embedded research taxonomy in the literature identified in Study 1. Once the APAT categories and criteria were developed, incorporating the American Psychological Association, Council for Exceptional Children (CEC), and criteria agreed to by the Delphi group, the APAT draft was given to the previous seven participants who completed the Delphi Survey method.

The seven participants were asked to review the instructions and the four APAT research designs (i.e., correlational, experimental/quasi-experimental, qualitative, and single-subject) for content, clarity, and consistency. Copies of these individual taxonomies are located in Appendices E through H.

Suggested changes to content were considered if they were supported by the American Psychological Association guidelines, CEC indicators, or were considered to be general edits (i.e., spelling, grammar, or flow). Suggested changes meeting these parameters were then reviewed by a professor considered to be an expert for each of the four research designs. The participant who did not provide specific feedback expressed concerns with using a rubric to evaluate the quality of research. A summary of suggestions are provided in Table 11.

Table 11

Sample of Suggested Participant Changes

	Correlational	Experimental/ quasi-experimental	Qualitative	Single-subject
Participant 1	Inquired into the inclusion of IRB	Inquired into the inclusion of IRB	Inquired into the inclusion of IRB	Inquired into the inclusion of IRB
Participant 2	Grammar edits and needed clarification of confusing wording	Grammar	Grammar	Grammar edits and needed clarification of confusing wording
Participant 3	Suggested reordering criteria in associated Groups (i.e., all criteria about participants grouped together)	Suggested reordering criteria in associated groups (i.e., all criteria about participants grouped together) and a design specific edit	Suggested reordering criteria in associated groups (i.e., all criteria about participants grouped together) and a design specific edit	Suggested reordering criteria in associated groups (i.e., all criteria about participants grouped together), identified a missing bullet, and expressed addressing appropriate single subject design
Participant 4	Editorial correction and identified a missing bullet	Had editorial correction	Needed clarification of an indicator	Editorial correction
Participant 5	Specific technical edits pertaining to the research design	Specific technical edits pertaining to the research design	Had specific technical edits pertaining to the research design	Technical edits pertaining to the research design

A complete listing of all suggested changes from each participant is available on request. However an example of these specific suggested changes from the participants is word for word [*sic*] and can be viewed in Table 12.

Table 12

Sample Summary of Participant Suggestions

Participant	Research design	Questions/Comments for suggested change(s)
Participant 1	Correlational	(Show IRB approvals?)
Participant 2	Correlational	Sampling bullet here in sequence. Include this concept throughout level 1-3 Should sampling be earlier in sequence
Participant 3	Correlational	Literature lacks <u>direct</u> connection to the purpose of the study (include underlined)
Participant 4	Correlational	Inclusion and exclusion of reported information is thoroughly addressed UNCLEAR
Participant 5	Correlational	Findings are compared to prior research (Evaluated in the context of previous research?)
Participant 1	Experimental	(IRB?)
Participant 2	Experimental	APAT: Level of Recommendation (Part III) looks fine
Participant 3	Experimental	Must all indicators be met within each domain before the standard is met? What if all are met but one or two? (same comments for all other research designs).
Participant 4	Experimental	Grammar corrections suggested
Participant 5	Experimental	Inclusion and exclusion of reported information is thoroughly addressed (From results section?)
Participant 1	Qualitative	(IRB?)
Participant 2	Qualitative	Literature supports purpose and justification of the study Interview questions are within the purpose of the study <u>and answer the research question and hypotheses</u> (include underlined)

(Continued)

Participant 3	Qualitative	Results are coded and presented but is not that is clear and confidential (Do not understand this indicator-needs clarification)
Participant 4	Qualitative	Measures are do not about insuring and maintaining the accuracy of recording participant results including transcribing are not mentioned (suggested change)
Participant 5	Qualitative	Population is adequately represented and includes description of inclusion/exclusion <u>criteria (This seems like 2 tasks/criteria)</u>
Participant 1	Single-Subject	(IRB?)
Participant 2	Single-Subject	Include examples in methods section selection of appropriate single-subject design for the study (i.e., reversal alternating treatment designs, multiple baseline, changing criterion)
Participant 3	Single-Subject	Grammar corrections suggested
Participant 4	Single-Subject	Baseline data are described operationally and provide evidence for repeated measurement UNCLEAR
Participant 5	Single-Subject	Participant criteria are fully described to the extent that is replicable and addresses participant confidentiality (Is that the same as inclusion/exclusion?)

Pilot Study Phase IV

Based on the results from the previous three phases in Study 2, within this investigation, the categories, criteria, and format of the four research designs that comprise the APAT were developed. Participants in this pilot test were 10 Adapted Physical Activity doctoral students who were asked to evaluate and score a purposefully selected research article within each of the four research designs. A professor in Kinesiology who is an expert in statistics was requested to purposefully select articles that met the following criteria: (a) it is available in full text; (b) written in English; (c) had a search perimeter that included “Adapted Physical Activity;” and (d) was

conducted using one of the four research designs. References for the selected articles to be evaluated are located on page 64.

The doctoral students were asked to read the instructions (Appendix D) and articles provided, and then evaluate the articles using the APAT. The doctoral students were also asked to identify the Strength of Quality (SOQ) and Level of Recommendation (LOR), and given the opportunity to provide written feedback regarding the clarity of the instrument, ease of use, and functionality of the instrument. Each of the doctoral students was able to use the APAT to select the correct research design evaluation form from the four choices. All of the doctoral students were able to provide a research design evaluation score for each of the articles.

In reviewing the four research designs, there was some disparity in the scoring the SOQ and LOR. The following are the SOQ and LOR with highest percent agreement among the doctoral students a score of less than .70 should draw significant caution (McMillan & Schumacher, 2014).

1. Correlational study evaluations: (a) SOQ level 2 = 50% agreement and (b) LOR A = 67% agreement.
2. Experimental/quasi-experimental study evaluations: (a) SOQ level 2 = 83% agreement and (b) LOR A = 83% agreement.
3. Qualitative study evaluation: (a) SOQ level 2 = 67% agreement and (b) LOR C = 50% agreement.
4. Single subject study evaluation: (a) SOQ level 2 = 67% agreement and (b) LOR A = 67% agreement.

Study 3: Phase V

The purpose of this study was to identify and to determine the inter-rater reliability of the APAT. Study 3 comprised one phase.

Inter-Rater Reliability

The purpose of Study 3 was to determine the inter-rater reliability (IRR) of the categories and criteria developed in Study 2, Phase II. A total of four different Psychdata surveys were developed; one for each of the four research designs. The first four questions of each survey were designed to establish the minimum requirement for participation in the study: (a) could not have previously participated in prior research in the development of the APAT; (b) had to have a ranking of tenured assistant professor or higher or could be retired at that level or higher; (c) had to have completed masters level education, or higher, in an area in the profession of Adapted Physical Activity; and (d) had taught at a university in the United States or Canada. The number of participants who met the requirements and participated in this part of this investigation are presented in Table 13. Please note that additional data is available from the researcher on request.

Table 13

Participant Recruitment

	Correlational	Experimental	Qualitative	Single-Subject
Total participants recruited	113	114	108	112
Total participants responding	34	37	27	34
Total participants meeting requirements	15	17	12	10

Of the combined 54 participants meeting the minimum requirements, only 24 participants completed the survey (44%). The 25 participants consisted of: (a) 3 Tenured Assistant Professors, (b) 12 Tenured Associate Professors, (c) 7 Tenured Full Professors, and (d) 3 Retired Full Professor.

The combined number of data-based refereed publications as first author of the participating participants was: 12 published 0 to 3 articles; 2 published 4 to 6; 2 published 9 to 15; and 6 published 16 or more. Further, the combined number of data-based non refereed publication as first author was: 14 participants published 0 to 3 articles; 4 published 7 to 9 articles; 3 published 10 to 15 articles; and 6 published 16 or more articles.

Each participant was provided one article that was previously purposefully selected and used during pilot testing (Phase IV). Participants were asked to read the article and evaluate the SOQ and LOR concurrently. Participants were provided the same information provided to the participants in the pilot testing (p. 63). However this information was presented within Psychdata using skip logic. Skip logic allows the participant to see only one section of the APAT in sequential order. For example, the participant must address Level 1: Introduction Section. If the participant determined all criteria were, the participant would then view Level 2: Method Section. If, however the participant answered “no” to Level 1: Introduction Section, which indicates the criteria was not met, the participant would then be asked if the Level 2 criteria for the Introduction Section was met.

The IRR was calculated by using the following formula:

$$IRR = \frac{\text{agreements}}{(\text{agreements} + \text{disagreements})}$$

Correlational Design

Of the 15 participants who met the participant requirements, 5 elected to complete the taxonomy for a response rate of 0.33 considered an inadequate return rate (Babbie, 1990). Based on Figure 12, 50% of the participants agreed it was of “Strong Quality,” 33% agreed it was of “Moderate Quality,” and 17% agreed it was of “Weak Quality.” The IRR coefficient for QOE was 0.60.

When asked to evaluate the LOR, 80% agreed it earned a LOR of “A” and 20% agreed it earned a LOR of “C” (see Figure 13). The IRR coefficient for LOR was 0.80.

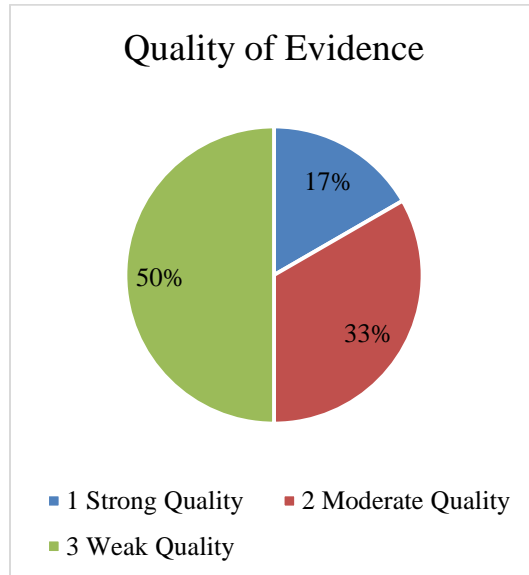


Figure 12. Percent of participants indicating strong, moderate, or weak SOQ for the correlational research design article.

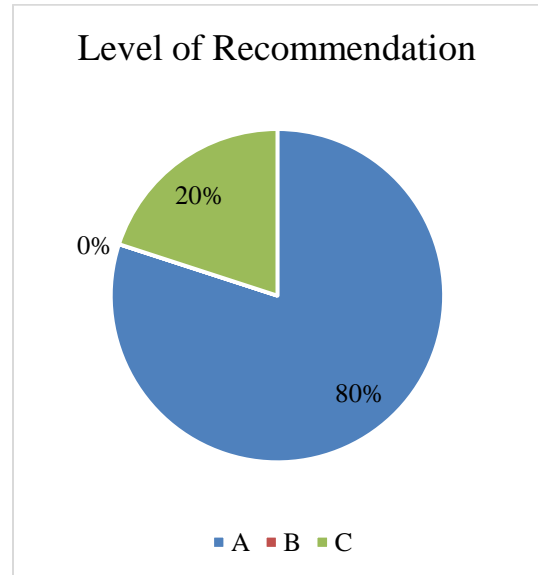


Figure 13. Percent of participants indicating Level A, B, or C LOR for the correlational research design article.

Experimental/Quasi-Experimental Design

Of the 17 participants who met the participant requirements, 7 elected to complete the taxonomy for a response rate of 0.41 which is considered an inadequate return rate (Babbie, 1990). Based on Figure 14, 57% of the participants agreed it was of “Strong Quality,” 14% agreed it was of “Moderate Quality,” and 29% agreed it was of “Weak Quality.” The IRR coefficient for QOE was 0.57. When asked to evaluate the LOR 72% agreed it earned a LOR “A,” 14% agreed it earned a LOR “B,” and 14% agreed it earned a LOR of “C” (see Figure 15). The IRR coefficient for LOR was 0.71.



Figure 14. Percent of participants indicating strong, moderate, or weak SOQ for the experimental/quasi-experimental research design article.

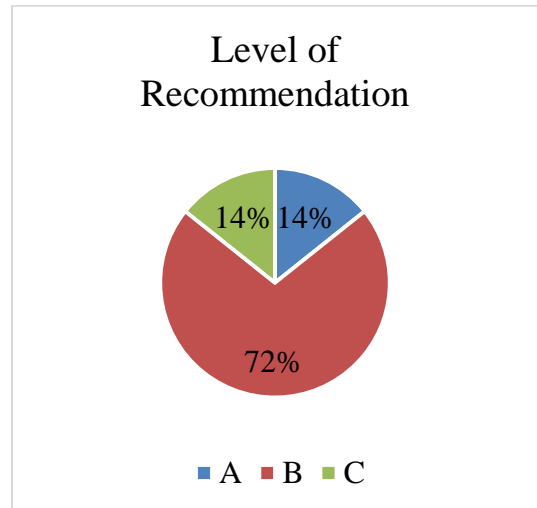


Figure 15. Percent of participants indicating Level A, B, or C LOR for the experimental/quasi-experimental research design article.

Qualitative Designs

The 12 participants who met the participant requirements, 9 elected to complete the taxonomy for a response rate of 0.75 considered a very good return rate (Babbie, 1990). Based on Figure 16, 56% of the participants agreed it was of “Strong Quality,” 22%

agreed it was of “Moderate Quality,” and 22% agreed it was of “Weak Quality.” The IRR coefficient for QOE was 0.55. When asked to evaluate the LOR 11% agreed it earned a LOR “A,” 67% agreed it earned a LOR “B,” and 22% agreed it earned a LOR of “C” (see Figure 17). The IRR coefficient for LOR was 0.66.

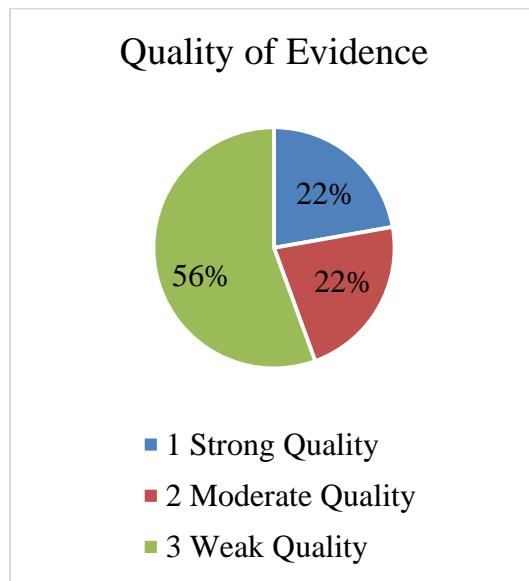


Figure 16. Percent of participants indicating strong, moderate, or weak SOQ for the qualitative research design article.

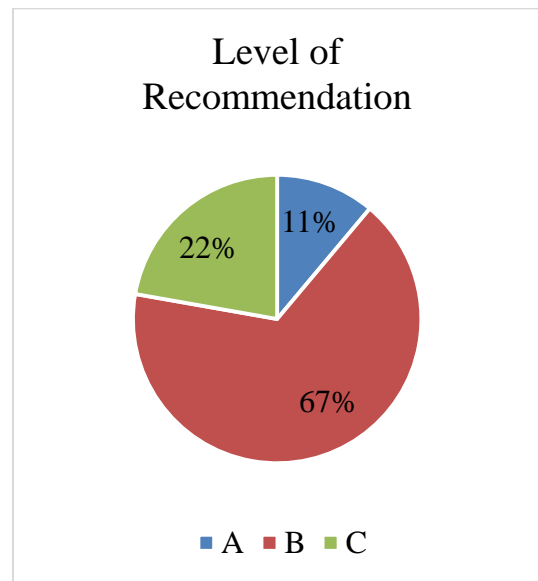


Figure 17. Percent of participants indicating Level A, B, or C LOR for the qualitative research design article.

Single-Subject Design

The 10 participants who met the participant requirements, 4 elected to complete the taxonomy for a response rate of 0.40 which is considered an inadequate return rate (Babbie, 1990). Based on Figure 18, 75% of the participants agreed it was of “Strong Quality” and 25% agreed it was of “Weak Quality.” The IRR coefficient for QOE was 0.75. When asked to evaluate the LOR 80% agreed it earned a LOR “A,” 25% agreed it

earned a LOR “B,” and 25% agreed it earned a LOR of “C” (see Figure 19). The IRR coefficient for LOR was 0.50.

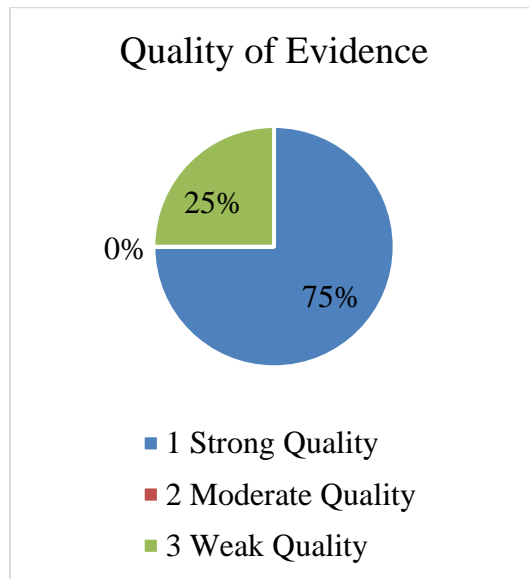


Figure 18. Percent of participants indicating strong, moderate, or weak SOQ for the single-subject research design article.

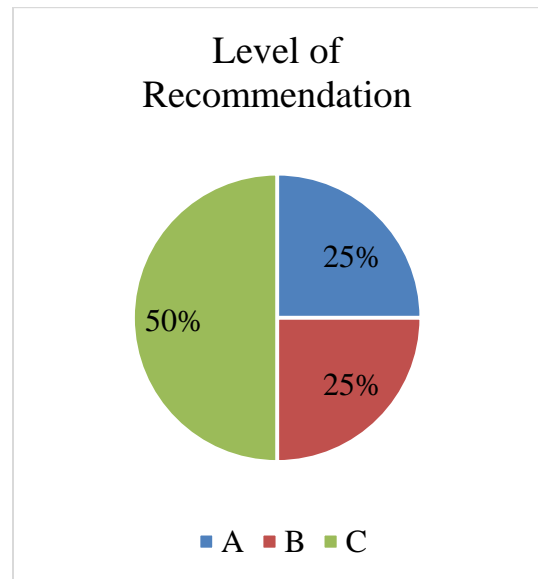


Figure 19. Percent of participants indicating Level A, B, or C LOR for the single-subject research design article.

It should be noted that, according to the standards identified by the CEC on IRR a score of 0.90 or higher is most preferred (Gersten, Fuchs, Coyne, Greenwood, & Innocenti, 2005) and a score of less than .70 should draw significant caution (McMillan & Schumacher, 2014).

CHAPTER V

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER STUDIES

Discussion

Since the mid 1980s, researchers have been expressing concerns related to the quality of research that has been conducted in Adapted Physical Activity (Lavay & Lasko-McCarthy, 1992; Reid & Stanish, 2003) and the need to increase the quality of research (Doll-Tepner & DePauw, 1996). The literature has become increasingly critical of the quality of published research in the profession of Adapted Physical Activity (Broadhead & Burton 1996; Porretta & Sherrill, 2005). Further, as productivity has increased, the trend in the use of evidence-based practices (EBP) has become of greater importance than in the past (Bouffard & Ried, 2012). In addition Hutzler (2012) stated that EBP today are the driving force behind quality research. A call for the use of scientifically based research was brought to the forefront in No Child Left Behind (NCLB, 2004). In response, the Council for Exceptional Children (CEC) began to identify what should be considered indicators for EBP (Classifying the State, 2008).

One method to evaluate the quality of EBP and level of recommendation of research is through the use of a research taxonomy. The first known taxonomy written in English was developed by the Canadian Task Force in 1979. The quality of research and the extent to which a treatment or practice should be recommended to a patient was identified

(Newman et al., 2007). Over the past 35 years more than 100 research taxonomies have been developed. By better identifying EBP (Weiss, 2004), the quality of research can be determined with more objectivity. Many of these taxonomies have been designed to focus on disease prevention and were developed in the professions of medicine, dentistry, and psychology (Baker et al., 2010; Kroke et al., 2004). Conversely the APAT was designed to focus primarily on the quality of published research by identify the strength of quality and level of recommendation that could be applied to research rather than being patient centered.

In the profession of Adapted Physical Activity, no research taxonomy could be located in the English language after an in-depth search of the literature. Based on this finding, Belanger and Colantonio (2011) conducted a survey to determine the need for a research taxonomy in the profession of Adapted Physical Activity. Based on the results of the survey, the participants, who were professors of Adapted Physical Activity, overwhelmingly supported the need for a research taxonomy.

Therefore, the purpose of this investigation was to develop a valid and reliable instrument to evaluate the quality of researcher for the profession of Adapted Physical Activity.

There are four purposes of this investigation:

1. To determine through a systematic search of literature to identify research taxonomies that exist in the English language, are used in any profession, and are embedded within published literature identified in an electronic database.

2. To identify categories and criteria embedded within the published research taxonomies.
3. To identify categories and criteria within the research articles and the criteria needed in an Adapted Physical Activity Taxonomy.
4. To determine the inter-rater reliability of the categories and criteria in the APAT that were developed.

A modified mixed-method meta-framework titled *Instrument Development and Fidelity* was used in this investigation (Onwuegbuzie et al., 2010). Five of the 10 phases were completed in this present investigation. The first five phases (I –V) involved justification and need for the instrument followed by the development and field-testing of the initial instrument. Phases VI through X consisted of validating and revising the instrument both quantitatively and qualitatively. The following are the major headings in this section: (a) systematic literature analysis, (b) identifying categories and criteria, (c) determining criteria needed for adapted physical activity taxonomy, (d) determining inter-rater reliability, (e) conclusions, and (d) recommendations for further studies.

Systematic Literature Analysis

No matter what the results of the searches and the present investigation, no research taxonomy could be located in the profession of Adapted Physical Activity. The three most complete published articles with findings related to taxonomies are as follows: (a) Shlutz and Foster (2011) and West et al. (2002) conducted their search electronically while; (b) Hillier et al. (2011) did not report how taxonomies located were identified. Table 14 provides the differences, similarities, and results of these three different

investigations. Of the taxonomies identified in this table there may be duplicate taxonomies between researchers. However, due to the lack of available, data duplicate taxonomies could not be determined. To date the researcher has been unable to identify a taxonomy that addressed the profession of Adapted Physical Activity.

Table 14

Summary of Taxonomy Development Findings

	Databases searched	Search parameters	Articles identified	Taxonomies identified
Present Investigation	10	12 phrases/terms	67	25
Hillier et al. (2011)	Not reported	Not reported	Not reported	9
Shurtz & Foster (2011)	14	4 phrases	Not reported	15
West et al. (2002)	2	4 questions	1,602	30

It should be noted while over 1,669 articles were identified and could be reviewed, one limitation to restricting the search to full text availability are articles available in print only. Of the three articles West et al. (2000) questioned the need for an intensive literature search. In support of West et al. the present investigator believed it is necessary to clearly determine and conclude there are no such taxonomies identified in the profession of Adapted Physical Activity. The discrepancy between West et al. and the number of articles identified in the present investigation may be the investigations of abstracts were only used. West et al. also conducted a manual search for articles, while

the present investigator focused on abstracts and articles that were available electronically in full text.

Identifying Categories and Criteria

In a review of the 25 taxonomies, the types of research designs were identified (e.g., mixed method, multiple research designs, qualitative, single-subject designs). In addition, the identification of the categories (i.e., major headings in research articles) and criteria (i.e., key statements) within each taxonomy were investigated.

As reported by West et al. (2002) taxonomies were not considered to be comprehensive enough in the ability to evaluate multiple designs. In contrast the expert evaluators in this present investigation identified the need to include multiple research designs within the APAT. Another finding was high quality could only be obtained by randomized control trials (RCT) in many of the current taxonomies including Grading of Recommendation Assessment, Development and Evaluation (GRADE; Baker et al., 2011). Similarly RCT and systematic reviews could receive a high ranking while all other studies could not earn a high ranking in Strength of Recommendation Taxonomy (SORT [Faggion, 2010]).

Participants in the current study did identify quantitative research as the most important and the research design used the most but did not identify it as a design that could only receive the highest ranking. Participants also identified the need to include mixed method design; at present only one of the 25 taxonomies included mixed method design (Clark, Burkett, & Stanko-Lopp, 2009). In addition, many taxonomies identified by West et al. required users to understand how to use various types of taxonomies to

evaluate multiple research designs. In the present investigation the instrument developed involved four taxonomies with a similar format. This supports West et al. in that the user will need to be familiar with more than one research design.

Within the research designs, 13 of the 25 taxonomies included categories (i.e., major headings in research articles) and criteria (i.e., key statements). Of those where categories and criteria were not identified, checklists (Moseley, Herbert, Sherrington, & Maher, 2002) and/or questions (Ware, & Gandek, 1998) were used. In other taxonomies there were no evaluation method identified (Summerfield, 1997). In contrast, in the present investigation the indicators identified by the expert participants from the CEC's and the American Psychological Association (2010) were to be included in the development of the APAT. Only 13 of the 25 taxonomies included categories (i.e., major headings in research articles), criteria (i.e., key statements), and focused on one or more research designs. It is important to note that in the taxonomy for this investigation, the same format for the four research designs differs because the criteria in the categories are based on the specific research design. However, how the Level of Recommendation is identified is the same for each taxonomy.

Determining Criteria Needed for Adapted Physical Activity Taxonomy

Experts in other professions than Adapted Physical Activity have developed research taxonomies according to protocols specific to the professions of pediatrics (Marcuse & Shiffman, 2004), speech and language (Gillam & Laing, 2006), nutrition (Woolf, 2006), dentistry (Faggion, 2010), music therapy (Abrams, 2010), and nursing (Pilcher & Bedford, 2011). The present investigator wanted to determine if a taxonomy with

specific categories and criteria needed for the profession of Adapted Physical Activity could be developed.

In their review of the profession of Adapted Physical Activity, Reid and Stanish (2003) noted there was a strong similarity between the profession of education and the medical profession and suggested the professions were cross disciplinary rather than independent. In contrast, participants overwhelmingly identified the general criteria of the American Psychological Association Standards (2010) and the Council for Exceptional Children Indicators (CEC, 2004) to be used for the development of the APAT criteria as opposed to criteria from the medical profession. This further demonstrates the need for continued development of a taxonomy in the profession of Adapted Physical Activity.

Taxonomies such as the one by the American Speech-Language-Hearing Association (<http://www.asha.org/academic/questions/Levels-of-Evidence/>, retrieved 2012), GRADE (Guyatt, et al., 2007), SORT (West, et al., 2002), Let Evidence Guide Every New Decision (Clark, Burkett, & Stanko-Lopp, 2009), and Scottish Intercollegiate Guidelines Network (<http://www.sign.ac.uk>, retrieved, 2012) to name a few, have all incorporated the use of categories within their taxonomies. Similarly the *Journal Article Reporting Standards* format from the American Psychological Association (2010) was used as the format for the categories for the APAT. After an opportunity to review the draft there were no comments or objections from the participants on the use of this format for the APAT. However, participants suggested other criteria inclusion in the APAT that did not blend into any of the categories. This criteria was then added in to the “Other Section”

where participants determined that the “overview of the study design” should be excluded while “solutions to challenges” should be included. The investigator developed the first APAT draft based on the format of SORT by infusing both the American Psychological Association standards and CEC indicators.

Determining Inter-Rater Reliability

The fourth purpose of this investigation was to determine the inter-rater reliability (IRR) of the APAT including the Strength of Quality (SOQ) and Level of Recommendation (LOR). As identified in the literature search SORT was the taxonomy and selected as the model for this investigation. The group developing SORT recommended that IRR be incorporated within future studies (West, et al., 2002). The present investigator statistically could not determine IRR because the APAT return rate was less than 30 evaluations per design that was the minimum needed for statistical analysis (p. 64). Therefore the findings were inconclusive (Faul et al., 2009).

Conclusions

Based on the review of the literature there are no other investigation related to evidence-based practice to evaluate the quality of research design and recommendations of research in the profession of Adapted Physical Activity. During the testing of the APAT a total of four articles were evaluated, one from each research designs. Three of the four articles were considered to be interventions. Previous taxonomies developed were to evaluate intervention based research. However, the APAT can evaluate the quality of an article even if it is not considered to be intervention.

The ability of the APAT to evaluate the quality of non intervention articles will meet the need for increasing the quality of research in the profession of Adapted Physical Activity. This is significant, because the conversation for quality research continues to be a concern in the profession (Sherrill & O’Conner, 1999) and the need to improve methods and the importance of statistical analysis is at the forefront (Sutlive & Ulrich, 1998). In order to provide a solution to this problem (Reid, Bouffard, & MacDonald, 2012; Sutlive & Ulrich, 1998) the APAT was developed. If the current APAT is recommended for use, each research design should be further evaluated (i.e., more experts, more on criteria) in order to ensure completeness, as well as, to determine the reliability of the APAT.

This is the first research investigation designed to develop an Adapted Physical Activity Taxonomy. An expert panel was surveyed and identified the criteria to be included within the APAT. As a result, this demonstrates a clear connection to the validity of the APAT. The next step in the development of the APAT will be to determine the IRR because of the low return rate in the present investigation.

Recommendations for Further Studies

To increase the effectiveness of further studies, it is recommended that an interest group be developed which would continue to investigate and refine the APAT. It is recommended that this group of highly qualified professionals in the profession of Adapted Physical Activity be committed to continually assess and reevaluate the APAT.

Recommendations include but are not limited to:

1. Consider implications of non tenured faculty in the profession of Adapted Physical Activity as possible participants for further IRR investigation. This study only recruited participants in Adapted Physical Education in the United States and Canada with tenure. It is recommended that PhD adjunct and assistant non tenured faculty in higher education be included in the future investigation.
2. Develop a database or a means of tracking evaluated articles to further investigate internal reliability. By establishing a database for researchers to catalog articles reviewed over time it may seek to identify potential weaknesses and pitfalls of the instrument.
3. Determine the possible implications for critically evaluating faculty and potential implications for advancement in the profession of Adapted Physical Activity in higher education. Because the instrument is still in development, institutions may use taxonomies that do not have an appropriate rating systems to their specific profession of inquire.

REFERENCES

- Abrams, B. (2010). Evidence-based music therapy practice: An integral understanding. *Journal of Music Therapy*, 47(4), 351-379. doi 10.1093/jmt/47.4.351
- Akkerman, S., Admiraal, W., Brekelmans, M., & Oost, H. (2008). Auditing quality of research in social sciences. *Quality & Quantity*, 42(2), 257-274. doi 10.1007/s11135-006-9044-4
- Allington, R. L. (2006). Reading lessons and federal policy making: An overview and introduction to the special issue. *Elementary School Journal*, 107(1), 3-15. doi 10.1086/509523
- Adapted Physical Activity Definition and Philosophy. (n.d.). Retrieved November 15, 2011 from <http://www.ifAdaptedPhysicalActivity.biz/?q=node/7>
- American Thoracic Society (2008). The GRADE Approach (Part 1 of 12) [Video file]. Retrieved from <https://www.youtube.com/watch?v=x6MlqC7157E>
- Babbie, E. 1990. *Survey research methods*, Wadsworth, Belmont, CA.
- Baker, A., Potter, J., Young, K., & Madan, I. (2010). A review of grading systems for evidence-based guidelines produced by medical specialties. *Clinical Medicine*, 10, 358-363. doi 10.7861/clinmedicine.10-4-358
- Baker, A., Potter, J., Young, K., & Madan, I. (2011). The applicability of grading systems for guidelines. *Journal of Evaluation in Clinical Practice*, 17, 758-762. doi:10.1111/j.1365-1365-2753.2011.01693.x

- Balshem, H., Helfand, M., Schünemann, H., Oxman, A., Kunz, R., & Brozek, J...Gunn E. V. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64, 401-406. doi:10.1016/j.jclinepi.2010.07.015
- Belanger, J., & Colantonio, S. (2011, March). *A survey of interest and need for an adapted physical activity taxonomy*. Poster session presented at the meeting of American Alliance for Health, Physical Education, Recreation, and Dance Student Symposium. San Diego, CA.
- Bouffard, M. (1993). The perils of averaging data in adapted physical activity research. *Adapted Physical Activity Quarterly*, 10(4), 371-391.
- Bouffard, M. (1997). Using old research ideas to study contemporary problems in adapted physical activity. *Measurement in Physical Education & Exercise Science*, 1(1), 71-87. doi 10.1207/s15327841mpee0101_5
- Bouffard, M., & Reid, G. (2012). The good, the bad, and the ugly of evidence-based practice. *Adapted Physical Activity Quarterly*, 29(1), 1-24.
- Bouffard, M., Strean, W. B., & Davis, W. E. (1998). Questioning our philosophical and methodological research assumptions: Psychological perspectives. *Adapted Physical Activity Quarterly*, 15(3), 250.
- Breslin, C., & Rudisill, M. (2011). The effect of visual supports on performance of the TGMD-2 for children with autism spectrum disorder. *Adapted Physical Activity Quarterly*, 28, 342-353.

- Brantlinger, E., Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children*, 71(2), 195-207. doi:10.1352/0047-6765@2005)42<92:IOEAWA>2.0.CO;2
- Brien, M., & Sveistrup, H. (2011). An intensive virtual reality program improves functional balance and mobility of adolescents with cerebral palsy. *Pediatric Physical Therapy*, 23(3), 258-266. doi 10.1097/PEP.0b013e318227ca0f
- Broadhead, G. (1990). Adapted physical activity: Terminology and concepts. In Doll-Tepper, G., Dahms, C., Doll, B., and von Selzam, H. (eds.). Adapted physical activity: An interdisciplinary approach. *Proceedings of the 7th International Symposium, Berlin, June 1989, New York, Springer-Verlag 1990, p. 3-9* United States.
- Broadhead, G., & Burton, A. (1996). The legacy of early adapted physical activity research. *Adapted Physical Activity Quarterly*, 13(2), 116-126.
- Canadian Task Force on the Periodic Health Examination. (1979). The periodic health examination. *Canadian Medical Association Journal*, 121, 1193-1252.
- Christina, R. W. (1989). Whatever happened to applied research in motor learning? In Skinner, J.S. et al. (ed.), *Future directions in exercise and sport science research*, Champaign, Ill., Human Kinetics.
- Clark, E., Burkett, K., & Stanko-Lopp, D. (2009). Let Evidence Guide Every New Decision (LEGEND): An evidence evaluation system for point-of-care clinicians and guideline development teams. *Journal of evaluation in clinical practice*, 15(6), 1054-1060.

Classifying the state of evidence for special education professional practices: CEC

Practice study manual. In *Council for Exceptional Children*. Retrieved from
[http://www.cec.sped.org/Content/NavigationMenu/ProfessionalDevelopment
/ProfessionalStandards/Practice_Studies_Manual_1_25.pdf](http://www.cec.sped.org/Content/NavigationMenu/ProfessionalDevelopment/ProfessionalStandards/Practice_Studies_Manual_1_25.pdf)

Cochrane, A. (1989). Archie Cochrane in his own words. Selections arranged from his
1972 introduction to "Effectiveness and efficiency: Random reflections on the
health services." *Controlled Clinical Trials*, 10(4), 428-433.
doi: 2065/10.1016/0197-2456(89)90008-1

Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based
practices in special education. *Exceptional Children*, 75(3), 365-383.
Retrieved from
[http://www.cec.sped.org/Content/NavigationMenu/Publications2/Exceptional
Children/](http://www.cec.sped.org/Content/NavigationMenu/Publications2/ExceptionalChildren/)

DePauw, K. P., & Doll-Tepper, G. M. (1989). European perspectives on adapted
physical activity. *Adapted Physical Activity Quarterly*, 6(2), 95-99.

Doll-Tepper, G., & DePauw, K. P. (1996). Theory and practice of adapted physical
activity: Research perspectives. *Sport Science Review Champaign*, 5, 1-11.

Ebell, M., Siwek, J., Weiss, B., Woolf, S., Susman, J., Ewigman, B., & Bowman, M.
(2004). Strength of recommendation taxonomy (SORT): A patient-centered approach
to grading evidence in the medical literature. *American Family Physician*, 69(3),
548-556. doi:10.3122/jabfm.17.1.59

- Faggion, Jr., C. M. (2010). Grading the quality of evidence and the strength of recommendations in clinical dentistry: A critical review of 2 prominent approaches. *Journal of Evidence-Based Dental Practice, 10*, 50-54.
doi:10.1016/j.jebdp.2010.01.001
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). *Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses*. Behavior research methods, 41(4), 1149-1160. doi 10.3758/BRM.41.4.1149
- García, C., Alvarado, K., & Gaxiola, G. (2011). Grading recommendations in clinical practice guidelines: Randomised experimental evaluation of four different systems. *Archives of Disease in Childhood, 96*(8), 723-728. doi:10.1136/abc.2010.199307
- Geerts, W., Pineo, G., Heit, J., Bergqvist, D., Lassen, M., Colwell, C., & Ray, J. (2004). Prevention of venous thromboembolism. *Chest, 126*, 338S-400S.
doi:10.1378/chest.119.1_suppl.132S
- Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164. doi 10.1177/001440290507100202
- Gillam, S., & Laing, R. (2006). Clinical forum. Making evidence-based decisions about child language intervention in schools. *Language, Speech & Hearing Services in Schools, 37*(4), 304. doi:10.1044/0161-1461(2006/035)
- Greer, N., Mosser, G., Logan, G., & Halaas, G. (2000). A practical approach to evidence grading. *Joint Commission Journal on Quality Improvement, 26*(12), 700-712.

- Guyatt, G. (2007). An emerging consensus on grading recommendations? *Chinese Journal of Evidence-Based Medicine*, 7(1), 1-4. doi:10.1136/ebm.11.1.2-a
- Guyatt, G., Oxman, A., Akl, E., Kunz, R., Vist, Jan Brozek, G.,... Schünemann, H., (2011). GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64, 383-394. doi:10.1016/j.jclinepi.2010.04.026
- Guyatt, G., Oxman, A., Schünemann, H., Tugwell, P., & Knottnerus, A. (2011). GRADE guidelines: A new series of articles in the journal of clinical epidemiology. *Journal of Clinical Epidemiology*, 64(4), 380-382. doi:10.1016/j.jclinepi.2010.09.011
- Guyatt, G., Oxman, A., Vist, G., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *Chinese Journal of Evidence-Based Medicine*, 9(1), 8-11. doi:10.1136/bmj.39489.470347.AD
- Hahn, D. (2009). Importance of evidence grading for guideline implementation: The example of asthma. *Annals of Family Medicine*, 7(4), 364-369. doi:10.1370/afm.995
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015. doi:10.1046/j.1365-2648.2000.t01-1-01567.x10.1046/j.1365
- Hillier, S., Grimmer-Somers, K., Merlin, T., Middleton, P., Salisbury, J., Tooher, R., & Weston, A. (2011). FORM: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC medical research methodology*, 11(1), 23.

- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179. doi 10.1177/001440290507100203
- Hutzler, Y. (2012). Evidence-based practice and research: A challenge to the development of adapted physical activity. *Adapted Physical Activity Quarterly, 28*(3), 189-209.
- Hutzler, Y., & Sherrill, C. (2007). Defining adapted physical activity: International perspectives. *Adapted Physical Activity Quarterly, 24*(1), 1-20.
- Institute of Education Sciences (ED), W. C. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Retrieved from <http://www2.ed.gov/rschstat/research/pubs/rigorousetid/rigorousetid.pdf>
- Jackson, D. (2010). Influence of evidence-based physical activity instruction on interpersonal and institutional ecological structures of post adjudicated youth. Texas Woman's University. *ProQuest Dissertations and Theses*, Retrieved From <http://ezproxy.twu.edu:2048/login?url=http://search.proquest.com/docview/857922374?accountid=7102>
- Jansma, P., & Surburg, P. (1995). Ph.D. competency guidelines and adapted physical education professional preparation in the United States. *Adapted Physical Activity Quarterly, 12*(4), 307-322.

- Johnson, R., Onwuegbuzie, A., & Turner, L. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(112), 112-133.
doi:10.1177/1558689806298224
- Kassner, M., Reid, G., & MacDonald, C. (2012). Evidence-based practice: Quality indicator analysis of antecedent exercise in autism spectrum disorders. *Research in Autism Spectrum Disorders*. doi:10.1016/j.rasd.2012.02.001
- Klingner, J. K., & Boardman, A. G. (2011). Addressing the "research gap" in special education through mixed methods. *Learning Disability Quarterly, 34*(3), 208-218.
doi:10.1177/0731948711417559
- Kroke, A., Boeing, H., Rossnagel, K., & Willich, S. N. (2004). History of the concept of 'levels of evidence' and their current status in relation to primary prevention through lifestyle interventions. *Public Health Nutrition, 7* (2), 279-284.
doi:10.1079/PHN2003535
- Lavay, B., & Lasko-McCarthy, P. (1992). Adapted physical activity research: Issues and recommendations. *Adapted Physical Activity Quarterly, 9*(3), 189-196.
- Levin, J. R. (2004). Random thoughts on the (in)credibility of educational-psychological intervention research. *Educational Psychologist, 39*(3), 173-184.
doi:10.1207/s15326985ep3903_3
- Linstone, H. A., & Turoff, M. (2002). *The Delphi method: Techniques and applications*. Retrieved from <http://is.njit.edu/pubs/delphibook/delphibook.pdf>
- Marcuse, E. K., & Shiffman, R. N. (2004). Classifying recommendations for clinical practice guidelines. *Pediatrics, 114*(3), 874-877. doi:10.1542/peds.2004-1260

- McKibbin, K. (1998). Evidence-based practice. *Bulletin of the Medical Library Association*, 86(3), 396-401.
- McMillan, J. H., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. Pearson Higher Ed.
- Moseley, A. M., Herbert, R. D., Sherrington, C., & Maher, C. G. (2002). Evidence for physiotherapy practice: A survey of the physiotherapy evidence database (PEDro). *Australian Journal of Physiotherapy*, 48(1), 43-49.
- NVivo qualitative data analysis software; QSR International Pty Ltd. Version 10, 2012.
- Newman, M., Weyant, R., & Hujoel, P. (2007). JEBDP improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews. *Journal of Evidence-Based Dental Practice*, 7, 147-150.
doi:10.1016/j.jebdp.2007.09.014
- No Child Left Behind Act of 2001, 20 U.S.C. §1112, 2011.
- O'Connor, J., French, R., Sherrill, C., & Babcock, G. (2001). Scholarly productivity in adapted physical activity pedagogy: A bibliometric analysis. *Adapted Physical Activity Quarterly*, 18(4), 434-450.
- Odom, S., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K. (2004). Quality indicators for research in special education and guidelines for evidence-based practices: Executive summary. Retrieved September 29, 2008, www.uoregon.edu/grantmatters/pdf/DR/Exec_Summary.pdf

- Odom, S., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, 71(2), 137-148. Retrieved from http://www.cec.sped.org/Content/NavigationMenu/ProfessionalDevelopment/ProfessionalStandards/Final_ManuscriptScientific_Research_in_Special_Education_R3.pdf
- Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research instrument: An example, design considerations and applications. *Information & Management*, 42 (1), 15-29.
- Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*, 4(1), 56-78. doi:10.1177/1558689809355805
- Oxman, A. D. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, 328, 1490-1494. doi: 10.1111/j.1398-9995.2010.02530.x
- Pilcher, J., & Bedford, L. A. (2011). Hierarchies of evidence in education. *Journal of Continuing Education in Nursing*, 42(8), 371-377. doi10.3928/00220124-20110401-03
- Pitney, W., & Parker, J. (2009). Qualitative research in physical activity and the health professions. *Champaign, Ill., Human Kinetics*.
- Porretta, D., Kozub, F., & Lisboa, F. (2000). Documentary analysis of survey research in adapted physical activity: 1984-1998. *Adapted Physical Activity Quarterly*, 17(3), 286-296.

- Porretta, D., Nesbitt, J., & Labanowich, S. (1993). Terminology usage: A case for clarity. *Adapted Physical Activity Quarterly*, 10(2), 87-96.
- Porretta, D. L., & Sherrill, C. (2005). APAQ at twenty: A documentary analysis. *Adapted Physical Activity Quarterly*, 22(2), 119.
- Porretta, D., Surburg, P., & Jansma, P. (2002). Perceptions of adapted physical education graduates from selected universities on attainment of doctoral competencies. *Adapted Physical Activity Quarterly*, 19(4), 420-434.
- Portney, L., & Watkins, M. P. (2007). *Foundations of clinical research: Applications to practice*. Pearson/Prentice Hall.
- Publication Manual of the American Psychological Association*, 6th ed.
Washington, DC US: American Psychological Association. 2010.
- Reid, G. (1989). Ideas About motor behavior research with special populations. *Adapted Physical Activity Quarterly*, 6(1), 1-10.
- Reid, G. (1992). Editorial. *Adapted Physical Activity Quarterly*, 9(1), 1-4.
- Reid, G. (1993). Motor behavior and individuals with disabilities: Linking research and practice. *Adapted Physical Activity Quarterly*, 10(4), 359-370.
- Reid, G. (2000). Future directions of inquiry in adapted physical activity. *Quest*, 52(4), 369-381. doi 10.1080/00336297.2000.10491724
- Reid, G., Bouffard, M., & MacDonald, C. (2012). Creating evidence-based research in adapted physical activity. *Adapted Physical Activity Quarterly*, 29(2), 115-131.
- Reid, G., & Stanish, H. (2003). Professional and disciplinary status of adapted physical activity. *Adapted Physical Activity Quarterly*, 20(3), 213.

- Reynolds, R. P. (1984). Research in adapted physical activity: Some reflections and modest proposals. *Adapted Physical Activity Quarterly*, 1(4), 263-266.
- Rodgers, B. L., & Cowles, K. V. (1993). The qualitative research audit trail: A complex collection of documentation. *Research in Nursing & Health*, 16 (3), 219-226. doi DOI: 10.1002/nur.4770160309
- Scottish Intercollegiate Guidelines Network (SIGN). (n.d.). Retrieved November 7, 2014.
- Shephard, R. (1999). Postmodernism and adapted physical activity: A new gnostic heresy?. *Adapted Physical Activity Quarterly*, 16(4), 332-343.
- Sherrill, C. (1997). Past, present, future. *Adapted Physical Activity Quarterly*, 14(1), 1-7.
- Sherrill, C. (2004). *Adapted Physical Activity, Recreation, and Sport: Cross disciplinary and Lifespan*. (5th ed). McGraw-Hill College.
- Sherrill, C., & O'Connor, J. (1999). Guidelines for improving adapted physical activity research. *Adapted Physical Activity Quarterly*, 16(1), 1-8.
- Shurtz, S., & Foster, M. J. (2011). Developing and using a rubric for evaluating evidence-based medicine point-of-care tools. *Journal of the Medical Library Association*, 99(3), 247.
- Siwek, J., Weiss, B., Woolf, S., Susman, J., Ewigman, B., & Bowman, M. (2004). Strength of recommendation taxonomy (SORT): A patient-centered approach to grading evidence in the medical literature. *American Family Physician*, 69(3), 548-556. doi:10.3122/jabfm.17.1.59

- Somerfield, M.R. (1997). Recommended breast cancer surveillance guidelines. *Journal of Clinical Oncology*, 15(5), 2149-2156.
- Strean, W. B. (1998). Identifying what we take for granted in our research: Suggestions for assumption hunters. *Adapted Physical Activity Quarterly*, 15(3), 278-284.
- Stuart, G. W., Tondora, J., & Hoge, M. (2004). Evidence-based teaching practice: Implications for behavioral health. *Administration and Policy in Mental Health*, 32(2), 107-130. doi:10.1023/B:APIH.0000042743.11286.bc
- Sutlive, V., & Ulrich, D. (1998). Interpreting statistical significance and meaningfulness in adapted physical activity research. *Adapted Physical Activity Quarterly*, 15(2), 103-118.
- Thangaratinam, S., & Redman, C. (2005). The Delphi technique. *Education: Obstetrician & Gynaecologist*, 7(2), 120-125. doi:10.1576/toag.7.2.120.27071
- Thompson, B., Diamond, K., McWilliam, R., Snyder, P., & Snyder, S. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children*, 71(2), 181-194. doi 10.1177/001440290507100204
- Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2005). *Research methods in physical activity*. Champaign, Ill., Human Kinetics.
- Tseng, M. H., Chen, K. L., Shieh, J. Y., Lu, L., & Huang, C. Y. (2011). The determinants of daily function in children with cerebral palsy. *Research in developmental disabilities*, 32(1), 235-245. doi 10.1016/j.ridd.2010.09.024

- Tsiros, M., Coates, A., Howe, P., Grimshaw, P., & Buckley, J. (2011). Obesity: The new childhood disability? *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity*, 12(1), 26-36.
doi:10.1111/j.1467-789X.2009.00706.x
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2003). *Identifying and implementing education practices supported by rigorous evidence: A user friendly guide*. Retrieved from <http://www2.ed.gov/rschstat/research/pubs/rigorousetid/rigorousetid.pdf>
- Ware Jr, J. E., & Gandek, B. (1998). Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *Journal of clinical epidemiology*, 51(11), 903-912. doi 10.1016/S0895-4356(98)00081-X
- Weiss, B. D. (2004). SORT: Strength of recommendation taxonomy. *Family Medicine*, 36(2), 141-143. Retrieved from <http://www.stfm.org/fmhub/fm2004/February/Barry141.pdf>
- West, S., King, V., Carey, T., Lohr, K., McKoy, N., Sutton, S., & Lux, L. (2002). Systems to rate the strength of scientific evidence. *Evidence report/technology Assessment*, 47, 1-11. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK11930/>
- What system of levels of evidence should be used in communication sciences and disorders? (n.d.). Retrieved November 7, 2014.

- Wilski, M., Nadolska, A., Dowling, S., Mcconkey, R., & Hassan, D. (2012). Personal development of participants in Special Olympics Unified Sports teams. *University School of Physical Education in Wrocław University School of Physical Education in Poznań University School of Physical Education in Kraków*, 271.
- Willoughby, K. L., Dodd, K. J., Shields, N., & Foley, S. (2010). Efficacy of partial body weight–supported treadmill training compared with overground walking practice for children with cerebral palsy: A randomized controlled trial. *Archives of physical medicine and rehabilitation*, 91(3), 333-339. doi: 10.1016/j.apmr.2009.10.029
- Woolf, S. H. (2006). Weighing the evidence to formulate dietary guidelines. *Journal of the American College of Nutrition*, 25(3 SUPPL.), 277S-284S.
doi: 10.1080/07315724.2006.10719578
- Zhang, J., deLisle, L., & Chen, S. (2006). Analysis of AAHPERD research abstracts published under special populations from 1968 to 2004. *Adapted Physical Activity Quarterly*, 23(2), 203-217.

Appendix A

Strength of Recommendation Taxonomy Summary
of Adapted Physical Activity Articles

Author	SORT Score	Research or Opinion Based	Summary
Belanger & Colantonio, 2012	L3	R	Survey study of professor's opinion on the need of a taxonomy for the field of APA.
Bouffard, 1993	L3	O	Opinion based article on the need for developing stronger quantitative analysis research in APA and urges the use of person-by-treatment research practice.
Bouffard, 1997	L3	O	Opinion based article that reflects on the use of using of "old research ideas" to analyze "contemporary problems".
Bouffard & Reid, 2012	L3	O	Opinion based article reflecting on the need for the development of evidenced-based practice in APA.
Bouffard, Streat & Davis, 1998	L3	O	Opinion based article questioning the need for a philosophical approach to research in APA.
Broadhead & Burton, 1996	L3	O	Opinion based article that indicates a greater need for depth of literature to substantiate and asks the question about what is needed to develop stronger research in APA.
Broadhead, 1990	L3	O	Identifies key terminology related to the field of APA.
Christina, 1989	L3	O	Opinion based article proposing the use of a three level system for evaluating research in APA.
DePaw & Doll-Tepner, 1989	L3	O	Article provides a historical perspective on the field of APA and defines key terminology as it relates to the field.
Doll-Tepner & DePaw, 1996	L3	O	A review of the history of research in the field of APA.
Hutzler, 2012	L3	R	Identifies a need for levels of evidence in APA.

Jansma & Surburg, 1995	L3	R	Through the use of survey Ph.D. competence for APA were developed.
Kasner, Reid, & MacDonald, 2012	L3	R	Experimental research design incorporating the use of CEC as quality indicators for APA.
Lavay & Lasko-McCarthy, 1992	L3	O	Suggests standards for use when reflecting on research in APA and address difficulties in conducting qualitative research in APA.
O'Connor, French, Sherrill, & Babcock, 2001	L3	R	Through a bibliometric analysis researchers identify the rate of publication in the field of APA.
Porretta & Sherrill, 2005	L3	O	Identifies key issues in research and how to develop a stronger research design and addresses issues in statistical interpretation.
Porretta, Kozub, & Lisboa, 2000	L3	R	A systematic review of survey research in APA.
Porretta, Surburg, & Jansma, 2002	L3	R	Survey of doctoral students regarding research awareness.
Reid & Stanish, 2003	L3	R	Systematic research that develops a foundational argument for tying APA to the medical field and borrowing taxonomy and incorporating quality indicators from the field of education.
Reid, 1993	L3	O	Identifies key components significant to motor development research.
Reid, 1989	L3	O	Opinion article regarding initial concepts of evaluation in APA research.
Reid, 2000	L3	O	Challenges the view of research by asking researchers to evaluating a researches philosophical approach.

Reid, Bouffard, & MacDonald, 2012	L3	O	Contends the development of evidence-based practice in APA can occur by asking five critical questions.
Reynolds, 1984	L3	O	Sets a challenge for researchers in the field of APA and identifies that much of the research in the field at that time was about psychological and service delivery issues.
Shephard, 1999	L3	O	Provides a critical appraisal of quantitative research in the field of APA.
Sherrill & O'Connor, 1999	L3	O	Identifies indicators that would potentially develop strong research in the field of APA.
Sherrill, 1997	L3	O	Opinion article identifying the need for research in APA to be theory driven.
Strean, 1998	L3	O	Identifies assumptions regarding research in the field of APA.
Sutlive & Ulrich, 1998	L3	O	Opinion article provides recommendations for research in APA suggesting more emphasis needs to be placed on statistical analysis.
Zhang, deLisle, & Chen, 2006, L3	L3	O	Literature review demonstrating an increase in non-experimental design; and raises the argument for including standards for other research designs for the field of APA.

Note: R = Research; O = Opinion based; APA = Adapted physical activity

Appendix B

First Round of Delphi Survey Documents and Questions

Description of Key Terms

Adapted physical activity (APA) - The cross-disciplinary body of knowledge directed toward the identification and solution of individual differences in physical activity. It is a service delivery profession and an academic field of study that supports an attitude of acceptance of individual differences, advocates access to activity lifestyles and sport, and promotes innovation and cooperative service delivery programs and empowerment systems. Adapted physical activity includes, but is not limited to, physical education, sport recreation, dance and creative arts, nutrition, medicine, and rehabilitation (International Federation of Adapted Physical Activity, retrieved 2012).

Criteria – The criteria may include national or local standards, quality indicators, or other levels that can be used to measure or identify the quality of research.

Level of evidence – The “level of evidence refers to the establishment of a hierarchy of study designs based on the ability of the design to protect against bias” (<http://www.asha.org/members/ebp/assessing.htm> retrieved on 10/17/2012).

Strength of recommendation – “The strength of a recommendation indicates the extent to which we [researchers] can be confident that adherence to the recommendation will do more good than harm” (Oxman, 2004, p. 1490).

Research designs – The design may include but are not limited to case study, experimental, quasi-experimental, single-subject, mixed method, qualitative, or other (Portney & Watkins, 2007).

Taxonomy – The instrument used to evaluate the quality of research and provide an evaluation score to represent the level of quality.

Adapted Physical Activity Taxonomy Development Delphi: Round 1

1. First and Last Name:
2. Check the box below indicating the total number of years as *Associate Professor or higher*.
☐ 0-3 years ☐ 3-6 years ☐ 6-9 years ☐ 9-15 years ☐ 15 or more
3. Please select your current academic rank
4. Select the number of *data based* refereed publications as first author.
☐ 0-3 ☐ 3-6 ☐ 6-9 ☐ 9-15 ☐ 16 or more
5. Select the number of *non-data* refereed publication as first author.
☐ 0-3 ☐ 4-6 ☐ 7-9 ☐ 10-15 ☐ 16 or more
6. Please *rank* which criteria for research standards should be used to develop an Adapted Physical Activity research taxonomy? (Select from the drop down boxes below: 1 indicating the most liked and 5 being least liked)

American Psychological Association Standards (2010)

Council for Exceptional Children (CEC) key quality indicators (Odom, S., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K., 2004 & 2005)

Grading of recommendations, assessment, development, and evaluation (Guyatt, Oxman, Scünemann, Tugwell, & Knottnerus, 2007)

PhD Competency Guidelines (Jansma & Surburg, 1995; see page 4)

Strength of Recommendation Taxonomy (Ebell, M., Siwek, J., Weiss, B., Woolf, S., Susman, J., Ewigman, B., & Bowman, M., 2004)

Other, please specify:

7. Please rank which research designs in APA is *most important* to you as a researcher. 1 indicating the one that is most important to 7 being least important) [Portney & Watkins, 2007].

Most Important

Mixed method
Qualitative
Quantitative
Single-subject
Correlational
Quasi experimental
Experimental

8. Please rank which research design you *use most as a researcher*: (Select from the drop down boxes below: 1 indicating the one that is most important to 7 being least used) [Portney & Watkins, 2007].

Used Most

Mixed method
Qualitative
Quantitative
Single-subject
Correlational
Quasi experimental
Experimental

9. Identified below are general headings in published research articles. Briefly describe what you would expect to read in the following sections.
- Introduction (Supporting studies, theory development, etc).
 - Method
 - Results
 - Discussion
 - Other, if so please specify:

10. It is generally agreed that the “Level of Evidence” is the foundation of a research taxonomy. But the question is, should the “strength of a recommendation indicate the extent to which we [researchers] can be confident that the recommendation will do more good than harm...?” (Oxman, 2004, p. 1490). Refer to page 1 for terminology. Please check the appropriate box:

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Agree
- ☐ Somewhat disagree
- ☐ Strongly disagree

11. What do you consider before you develop a research design to help address the level of evidence? (e.g., a study design, statistical analysis, participant selection). Briefly describe:

Doctoral Competency Guidelines for Research in Adapted Physical Education

(Jansma & Surburg, 1995)

1.0 Researcher

- 1.1 Understand the framework underlying scientific theory, reasoning, and research.
- 1.2 Identify and comprehend current research issues in adapted physical education, special education, and supporting areas.
- 1.3 Formulate relevant research questions, including research design, with recognition of the inherent limitations/advantages.
- 1.4 Acquire research-related skills in data acquisition and descriptive data formulation.
- 1.5 Apply research design theory to parametric, nonparametric, correlational, factor analysis, and regression research.
- 1.6 Understand procedures of behavioral, qualitative, and survey research.
- 1.7 Select, apply, and interpret statistics appropriate for selected research designs.
- 1.8 Demonstrate the ability to design research projects that involve the unique problems of special populations.
- 1.9 Demonstrate computer literacy in the areas of database management, data reduction, and data analysis.
- 1.10 Demonstrate the ability to analyze different types of data sets.
- 1.11 Write the results of a research problem in publishable form which includes a problem statement, hypothesis, review of literature, methods and procedures, research design, data collection, data analysis, and discussion.
- 1.12 Critique, in a meaningful manner, research designs described in refereed journals in terms of the appropriateness of research problems.
- 1.13 Demonstrate the ability to research and prepare proposals for funding research in adapted physical education.
- 1.14 Demonstrate the capacity to be a valuable contributor to a research project as an associate researcher.
- 1.15 Conduct a research project involving special populations.
- 1.16 Conduct computer searches of relevant research literature from extant databases such as ERIC, CD-ROM, and MED-LINE.
- 1.17 Present a data-based study at a state, regional, national, or international conference.
- 1.18 Maintain an awareness of relevant issues and trends in research.

Appendix C
Second Round of Delphi Survey Questions

Adapted Physical Activity Taxonomy Development: Delphi (Round 2)

1. First and Last Name:

2. Based on the responses from Survey 1, the following are the 3 research designs that were considered the *least important*: (a) Correlation, (b) Single-Subject, and (c) Qualitative. To verify that these designs should be **EXCLUDED** from the APA taxonomy, please check the appropriate box for each of the designs.

Correlation	Single-Subject	Qualitative
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

3. Based on the responses from Survey 1, the following is information that should be **INCLUDED** within the **Introductory Section** of research articles. Please check the appropriate box below:

Statement of the Problem/Purpose	Overview of the Study Design	Supporting Literature	Rational/Challenges/Justification to the Study
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

Theory or Conceptual Model of Development	Summary of the Knowns and Unknowns	Research Questions and Hypothesis	Solutions to Challenges
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

4. Based on the responses from Survey 1, the following is information that should be **INCLUDED** within the **Method Section** of research articles. Please check the appropriate box below:

Identification of the Participants (e.g., Subject Recruitment)	Data Collection, Treatment, and Analysis	Description of the Study Design that is Replicable (e.g., Procedures, Instrument, Setting)
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

Appropriateness of Answering Research Question (e.g., Align with dependent variable)	Threats to Internal and External Validity
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

5. Based on the responses from Survey 1, the following is information that should be **INCLUDED** within the **Results Section** of research articles. Please check the appropriate box below:

Provide Data	Description and Analyses of Raw Data (e.g., Tables, Figures)	Statistical Treatment of Data (e.g., Aligned with Methods, Ensure Replication)	Results Address the Research Question
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

6. Based on the responses from Survey 1, the following is information that should be **INCLUDED** within the **Discussion Section** of research articles. Please check the appropriate box below:

Results Address the Research Question	Compare the Findings to Prior Research	Identification of Limitations of the Study	Recommendation for Future Studies
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

7. Based on the responses from Survey 1, the following is information that should be **INCLUDED** as identified in the “**Other**” section of the survey. Please check the appropriate box below:

References	Appendices
<input type="checkbox"/> Exclude <input type="checkbox"/> Include	<input type="checkbox"/> Exclude <input type="checkbox"/> Include

8. Based on Survey 1 (Question #11), when developing a study design, the following items listed below are to be considered for the **Level of Evidence**. Level of evidence is defined as level of evidence refers to the establishment of a hierarchy of study designs based on the ability of the design to protect against bias” (<http://www.asha.org/members/ebp/assessing.htm> retrieved on 10/17/2012).

Please select from the drop down boxes below: 1 indicating the most important to 4 being least important please use a number only once.

Pilot study was conducted
 Clearly identified research question/problem – hypotheses to?
 Research design that appropriately fits the study
 Validation/pros and cons of various methods

Appendix D
Third Round of Delphi Survey Questions

Adapted Physical Activity Taxonomy Development: Delphi (Round 3)

12. First and Last Name:

13. In Round 2 of the survey, 70% agreement or disagreement was not met on 2 statements. Below are the 2 statements that need to be reconsidered in a Likert Type format. Please indicate to what extent you agree that an “*Overview of the Study Design*” and “*Solutions to Challenges*” should be **INCLUDED** within the **Introductory Section** of research articles. Please check the appropriate box below:

“Overview of the Study Design”	“Solutions to Challenges”
<input type="checkbox"/> Strongly agree <input type="checkbox"/> Somewhat agree <input type="checkbox"/> Agree <input type="checkbox"/> Somewhat disagree <input type="checkbox"/> Strongly disagree	<input type="checkbox"/> Strongly agree <input type="checkbox"/> Somewhat agree <input type="checkbox"/> Agree <input type="checkbox"/> Somewhat disagree <input type="checkbox"/> Strongly disagree

14. In Round 2 of the survey, 70% agreement or disagreement was not met on 1 statement. Below is 1 statement that needs to be reconsidered in a Likert Type format. Please indicate to what extent you agree that “*Appendices*” should be **INCLUDED** as identified in the “**Other**” section in research articles. Please check the appropriate box below:

“Appendices”
<input type="checkbox"/> Strongly agree <input type="checkbox"/> Somewhat agree <input type="checkbox"/> Agree <input type="checkbox"/> Somewhat disagree <input type="checkbox"/> Strongly disagree

Appendix E

Adapted Physical Activity Taxonomy Instruction Sheet

Instructions for how to use the Adapted Physical Activity Taxonomy

The APAT is comprised of two parts: Part 1 - Review for Quality of Study and Part 2 - Review for Level of Recommendation. Instructions for how to use the APAT to evaluate a research study is listed below.

Part I: Review for Quality of Study:

- Review each section and identify the *Quality of Study* related to the manuscript/article that you are reviewing.
- Identify the research design used the manuscript/article. If an article is opinion based and not identified as research based, it would receive a Level 3 ranking. If not opinion based, select the research design, then refer to the corresponding taxonomy (i.e., single-subject, correlation).
- If all the indicators for the specific research design are *not* met in Level 1, the reviewer proceeds to Level 2.
- Each section heading is to be evaluated until all sections have been reviewed. For example, a reviewer begins with the **Introductory Section**, if all indicators are met the reviewer moves to the **Method Section**, and begins the same process. For a manuscript/article to be classified as a Level 1 it must meet *all* the indicators for the research design in all domains.
- If a reviewer evaluates a manuscript/article and ALL indicators in the **Method Section** in Level 1 are not met such as “research design appropriately aligning with the hypothesis(es),” the reviewer would drop down to the Level 2 and determine if the “research design somewhat aligns with the hypothesis(es).” If it meets these indicators and all other indicators within Level 2 the reviewer would continue across the Level 2 indicators. If ALL indicators in Level 2 are not met, the reviewer would proceed to Level 3 for ALL indicators. If ALL indicators in Level 3 are not met, the study would be considered the lowest level, Level 3.
- Ranking of a manuscript/article related to the Quality of the study, is based on the lowest level of identified indicators across ALL sections.
- Once the reviewer completes Part I, move to Part 2 to determine the Level of Recommendation.

Part II: Review for Level of Recommendation

- Start at the top of Part 2 of the form and read the first box on the left. Follow the arrow that indicates “yes” it meets the indicators or “no” it does not meet the indicators.
- Continue to answer the questions until you reach a level of recommendation of A, B, or C.

Appendix F

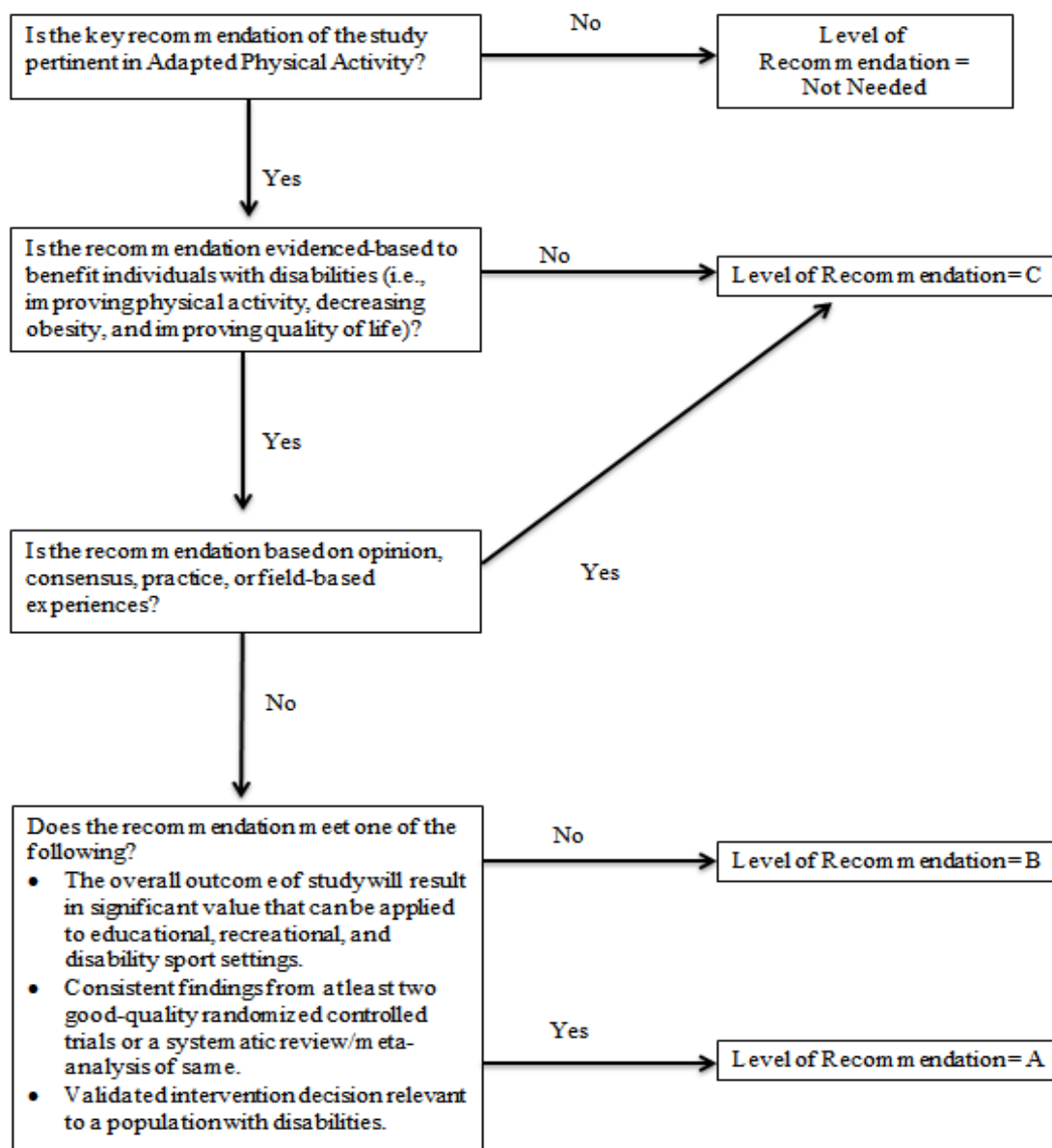
Final Adapted Physical Activity Taxonomy: Correlational Design

Part I: Quality of the Study					
APAT: Correlational Design Based on the Strength of Recommendation Format					
Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 1 (Strong)	Indicators: <input type="checkbox"/> Hypothesis(es)/ research question(s) is/are clearly stated <input type="checkbox"/> Purpose of the study clearly align with hypothesis(es)/ research questions(s) <input type="checkbox"/> If appropriate there is a clear application to theory and/or conceptual model <input type="checkbox"/> Significance and need for the study are clearly demonstrated <input type="checkbox"/> Solutions to the challenges are provided <input type="checkbox"/> Literature supports the purpose and justification of the study	Indicators: <input type="checkbox"/> Research design appropriately aligns with the hypothesis(es)/research question(s) <input type="checkbox"/> Research instrument(s) have currently been validated and show reliability within the target population <input type="checkbox"/> Appropriate measures are used to control for participant and researcher bias <input type="checkbox"/> Participants selected clearly reflect the intent of the study <input type="checkbox"/> Participant confidentiality is clearly addressed <input type="checkbox"/> Sampling technique(s) is/are described in depth <input type="checkbox"/> Population is adequately represented and includes description of inclusion and/or exclusion criteria <input type="checkbox"/> Researcher(s) provides sufficient information for replication of the study including a clear description of the setting <input type="checkbox"/> Intervention and/or conditions are thoroughly explained and documented	Indicators: <input type="checkbox"/> Percent agreement between observers is $\geq 90\%$ or coefficient r is ≥ 0.7 <input type="checkbox"/> Analyses of raw data were clearly described <input type="checkbox"/> Effect size is provided <input type="checkbox"/> Confidence intervals are presented for reliability coefficients and sample statistics <input type="checkbox"/> Univariate measures are used appropriately <input type="checkbox"/> Reliability and validity interpretations are very detailed	Indicators: <input type="checkbox"/> Discussion of results clearly address the hypothesis(es)/research question(s) <input type="checkbox"/> Findings are compared to appropriate prior research <input type="checkbox"/> Limitations of the study are clearly identified <input type="checkbox"/> Recommendation for future studies are clear and detailed <input type="checkbox"/> Generalizability clearly addresses target population and other possible issues	Indicators: <input type="checkbox"/> Contains a complete listing of references pertinent to the study <input type="checkbox"/> Appropriate appendices are provided
	Meets standards, move to L1 Method Section	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 2 (Moderate)	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Hypothesis(es)/ research question(s) is/are present but not clear <input type="checkbox"/> Purpose of the study does not clearly align with hypothesis(es)/ research question(s) <input type="checkbox"/> If appropriate there is limited application to theory and/or conceptual model <input type="checkbox"/> Significance and need for the study are not clearly demonstrated <input type="checkbox"/> Some solutions to challenges are provided <input type="checkbox"/> Literature lacks direct connection to the purpose and justification of the study <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Research design somewhat aligns with the hypothesis(es)/research question(s) <input type="checkbox"/> Research instrument(s) does not clearly demonstrate current validity and reliability within the target population <input type="checkbox"/> Limited measures are used to control for participant and researcher bias <input type="checkbox"/> Participants who have been selected vaguely reflect the intended purpose of the study <input type="checkbox"/> Participant confidentiality is minimally addressed <input type="checkbox"/> Population is minimally represented and does not provide a clear description of inclusion and/or exclusion criteria <input type="checkbox"/> Sampling technique(s) is/are described but is not replicable and setting is adequately described <input type="checkbox"/> Researcher(s) provides adequate information for replication of study <input type="checkbox"/> Intervention and/or conditions are adequately explained and documented <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Percent agreement between observers is between 80% to 89%, or coefficient r is between 0.5 to 0.69 <input type="checkbox"/> Analyses of raw data were presented but description lack clarity <input type="checkbox"/> Effect size is inconsistently provided <input type="checkbox"/> Confidence intervals are partially presented for reliability coefficients and sample statistics <input type="checkbox"/> Univariate measures may or may not be used appropriately <input type="checkbox"/> Reliability and validity interpretations provide some detail <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Discussion of results does not clearly address the hypothesis(es)/research question(s) <input type="checkbox"/> Findings are present but not all are compared to prior research <input type="checkbox"/> Limitations are mentioned but not clearly identified <input type="checkbox"/> Recommendations for future studies are mentioned but not addressed in detail <input type="checkbox"/> Generalizability addresses target population and other possible issues but is limited <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Partial inclusion of references pertinent to the study <input type="checkbox"/> Appendices are incomplete <p>Indicator evaluation</p>

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 3 (Weak)	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Hypothesis(es)/ research question(s) are not present <input type="checkbox"/> Purpose of the study does not align with hypothesis(es)/ research question(s) <input type="checkbox"/> If appropriate there is no application to theory and/or conceptual model <input type="checkbox"/> Significance and need for the study is not identified or is opinion based <input type="checkbox"/> Solutions to challenges are not provided <input type="checkbox"/> Literature has no clear connection to the purpose and justification of the study 	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Research design does not align with the hypothesis(es)/research question(s) <input type="checkbox"/> Research instruments are not currently valid and reliable <input type="checkbox"/> No apparent measures are taken to control for participant and researcher bias <input type="checkbox"/> Participants who have been selected do not reflect the intended purpose of the study <input type="checkbox"/> Participant confidentiality is not addressed <input type="checkbox"/> Population is not represented and does not provide inclusion and/or exclusion criteria <input type="checkbox"/> Sampling technique(s) is/are not provided <input type="checkbox"/> Researcher(s) provides insufficient information for replication of study and setting is not addressed <input type="checkbox"/> Intervention and/or conditions are inadequately explained and/or documented 	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Percent agreement between observers is $\leq 79\%$, or coefficient r is ≤ 0.49 <input type="checkbox"/> Analyses of raw data were missing and/or not clear <input type="checkbox"/> Effect size is not provided <input type="checkbox"/> Confidence intervals are not presented for reliability coefficients and sample statistics <input type="checkbox"/> Univariate measures are not used correctly <input type="checkbox"/> Reliability and validity interpretations provide little to no detail 	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Discussion of results does not address hypothesis(es)/research question(s) <input type="checkbox"/> Findings are not compared to appropriate prior research <input type="checkbox"/> Limitations are not identified <input type="checkbox"/> Limited or no recommendations to future research presented <input type="checkbox"/> Generalizability does not address target population and other possible issues 	<p>Indicators:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Does not include references pertinent to the study <input type="checkbox"/> Appendices are not provided <p>Indicator evaluation</p>

APAT: Level of Recommendation (Part II)



Used and modified from Newman, M., Weyant, R., & Hujoel, P. (2007). JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews. *Journal of Evidence-Based Dental Practice*, 7, 147-150. doi:10.1016/j.jebdp.2007.09.014

Appendix G

Final Adapted Physical Activity Taxonomy:

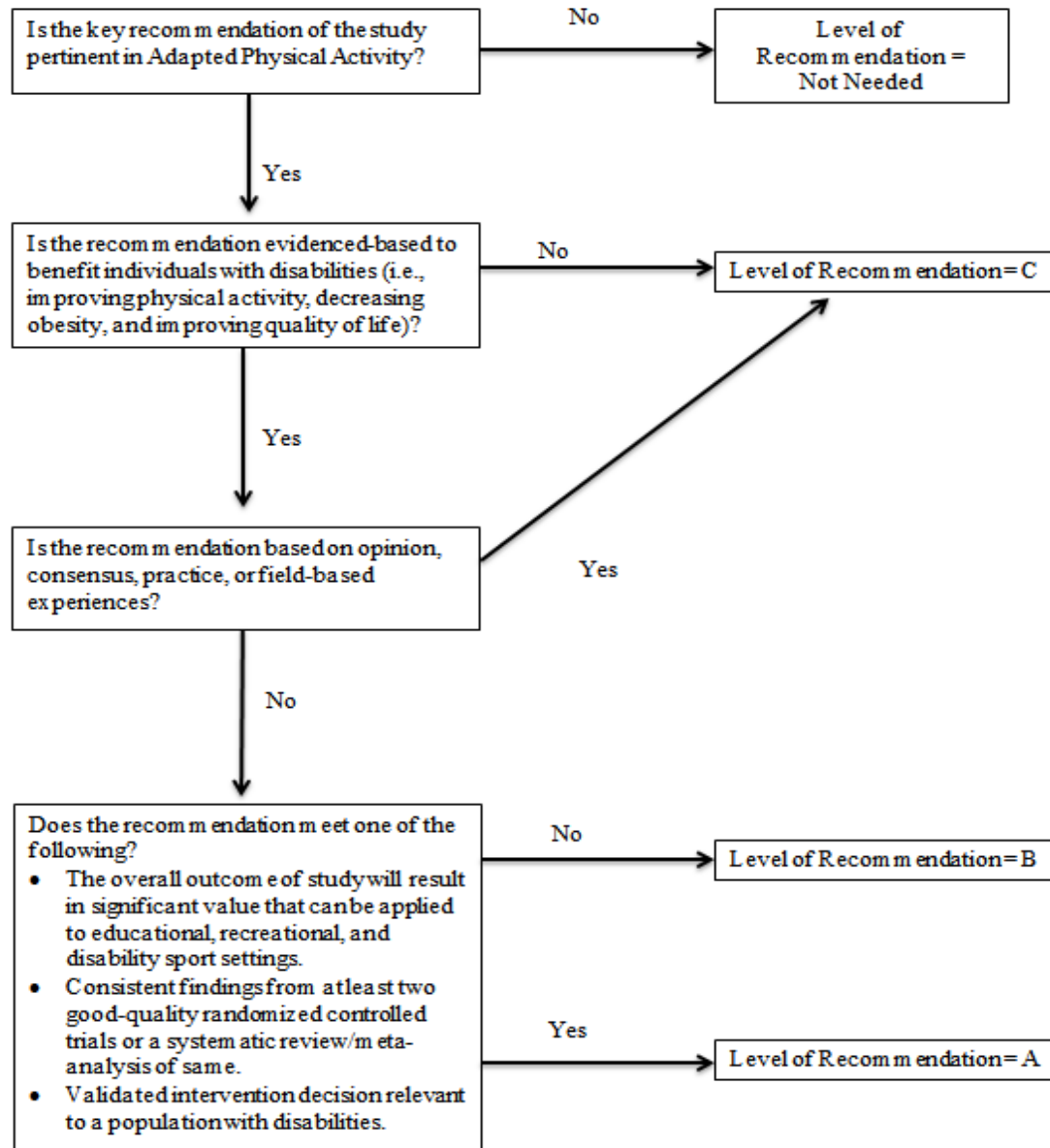
Experimental/Quasi-Experimental Design

Part I: Quality of the Study		APAT: Experimental/Quasi Experimental Design Based on the Strength of Recommendation Format			
Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 1 (Strong)	Indicators: <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are clearly stated Purpose of the study is clearly aligned with hypothesis(es)/ research question(s) If appropriate, there is a clear application to theory and/or a conceptual model Significance and need for the study is clearly demonstrated Solutions to the challenges are provided Literature supports purpose and justification of the study 	Indicators: <ul style="list-style-type: none"> Research design appropriately aligns with the hypothesis(es)/research question(s) Research instrument(s) have been currently validated and show reliability within the target population Appropriate measures are used to control for participant and researcher bias Participants selected, clearly reflect the intended study Population is adequately represented and includes description of inclusion/exclusion criteria Participant confidentiality is clearly addressed Participants characteristics are comparable across conditions Data collection was conducted throughout the course of the treatment Sampling technique(s) is/are described in depth Researcher(s) provide(s) sufficient information for replication of the study and includes description of the setting Treatment levels, intervention, and/or conditions are thoroughly explained and/or documented Fidelity is clearly described and assessed 	Indicators: <ul style="list-style-type: none"> Power is ≥ 0.8 Clear description and analyses of raw data Results address the hypothesis(es)/ research question(s) Effect size is provided Reliability and validity interpretations are very detailed 	Indicators: <ul style="list-style-type: none"> Discussion of results clearly address the hypothesis(es)/research question(s) Findings are compared to appropriate prior research Limitation of the study are clearly identified Recommendation for future studies are clear and detailed Generalizability addresses target population and other possible issues 	Indicators: <ul style="list-style-type: none"> Contains a complete listing of references pertinent to the study Appendices are provided
	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 2 (Moderate)	<p>Indicators:</p> <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are present but not clear Purpose of the study does not clearly align with hypothesis(es)/ research question(s) If appropriate, there is limited application to theory and/or conceptual a model Significance and need for the study is not clearly demonstrated Some solutions to challenges are provided Literature lacks direct connection to the purpose and justification of the study <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Research design somewhat aligns with the hypothesis(es)/research question(s) Research instrument(s) does not clearly demonstrate current validity and reliability in target population Limited measures are used to control for participant and researcher bias Participants selected vaguely reflect the intended purpose of the study Population is underrepresents and may not provide a description of inclusion/exclusion criteria Participant confidentiality is present but not clear Participants characteristics have limited comparison across conditions Data collection was conducted throughout the course of the treatment Sampling technique(s) is/are described but is not replicable Researcher(s) provide adequate information for replication of study and setting is adequately described Treatment levels, intervention, and/or conditions are adequately explained and documented Fidelity is not fully described and assessed <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Power is between 0.7 to 0.79 Analyses of raw data were presented but description lacks clarity Not all results align with hypothesis(es)/ research questions Effect size is inconsistently provided Reliability and validity interpretation lack detail <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Discussion of results does not clearly address the hypothesis(es)/research question(s) Findings are present but not all are connected to appropriate prior research Limitations are mentioned but not clearly identified Recommendation are mentioned but not addressed in detail Generalizability addresses target population and other possible issues but is limited <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Partial inclusion of references pertinent to the study Appendices are incomplete <p>Indicator evaluation</p>

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 3 (Weak)	<p>Indicators:</p> <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are not present Purpose of the study does not align with hypothesis(es)/ research question(s) If appropriate there is no application to theory and/or conceptual model Significance and need for the study is not identified or is opinion based Solutions to challenges are not provided Literature has no clear connection to the purpose and justification of the study 	<p>Indicators:</p> <ul style="list-style-type: none"> Research design does not align with the hypothesis(es)/research question(s) Research instrument(s) does not demonstrate current validity and reliability No apparent measures are taken to control for participant bias Participants selected do not reflect the intended purpose of the study Participant selection does not align with study and does not include a description of inclusion/exclusion criteria Participant confidentiality is not addressed Participant characteristics are not comparable across conditions Data collection was not conducted throughout the course of the treatment Sampling technique(s) is/are not provided Researcher(s) provide insufficient information for replication of study and setting is not addressed Treatment level, intervention, and/or conditions are inadequately explained and/or documented Fidelity is not described and assessed 	<p>Indicators:</p> <ul style="list-style-type: none"> Power is ≤ 0.69 Raw data were missing and/or not clear Results do not align with hypothesis(es)/ research questions Effect size is not provided Reliability and validity interpretation are unclear and without detail 	<p>Indicators:</p> <ul style="list-style-type: none"> Discussion does not clearly address the hypothesis(es)/research question(s) Findings are not connected to appropriate prior research Limitations are not identified Limited or no recommendations to future research are presented Generalizability does not addresses target population and other possible issues 	<p>Indicators:</p> <ul style="list-style-type: none"> Does not include references pertinent to the study Appendices are not provided <p>Indicator evaluation</p>

APAT: Level of Recommendation (Part II)



Used and modified from Newman, M., Weyant, R., & Hujoel, P. (2007). JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews. *Journal of Evidence-Based Dental Practice*, 7, 147-150. doi:10.1016/j.jebdp.2007.09.014

Appendix H

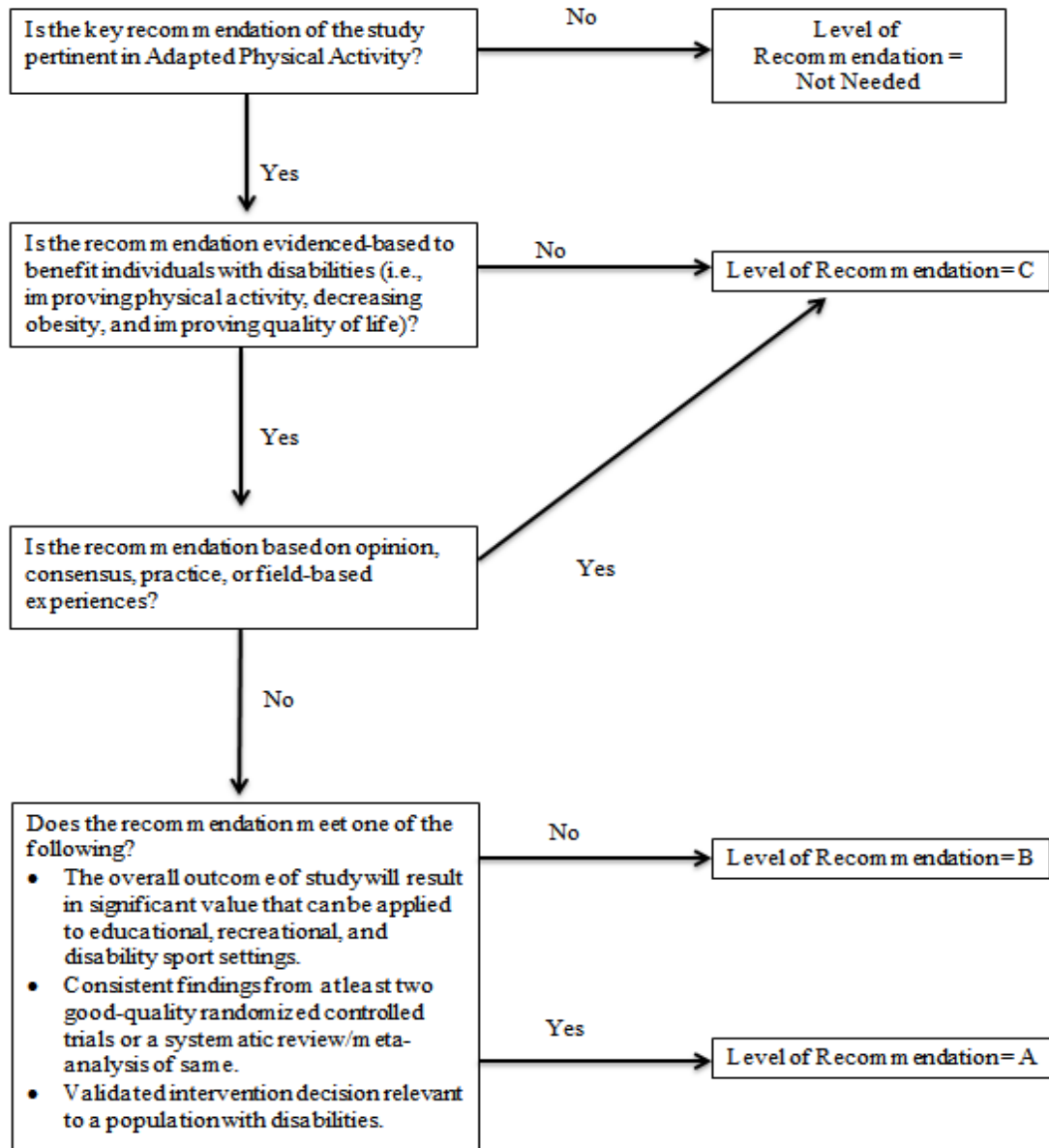
Final Adapted Physical Activity Taxonomy: Qualitative Design

Part I: Quality of the Study		APAT: Qualitative Design Based on the Strength of Recommendation Format			
Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 1 (Strong)	Indicators: <ul style="list-style-type: none"> □ Research question(s) is/are clearly stated □ Purpose of the study clearly aligns with the research question(s) □ Clear application to theory and/or conceptual model □ Significance and need for the study is clearly demonstrated □ Solutions to the challenges are provided □ Literature supports the purpose and justification of the study 	Indicators: <ul style="list-style-type: none"> □ Research design appropriately aligns with the research question(s) □ Appropriate measures are used to control for participant and researcher bias □ Participants selected, clearly reflect the intended study □ Population includes full description of inclusion and exclusion criteria □ Participant criteria are fully described to the extent that it is replicable □ Participant confidentiality is clearly addressed □ Data collection was conducted throughout the course of the treatment to substantiate trustworthiness; if appropriate, preliminary data measurements are obtained □ Sampling technique(s) is/are clearly described in depth □ Researcher(s) provide sufficient information for replication of the study and includes description of the setting □ Treatment levels are thoroughly explained and documented □ Interview questions are within the purpose of the study and answer the research question □ Measures are taken to explain the accuracy of the description of the treatment of collected data □ If appropriate documents/artifacts are meaningful, accurately described, and cited appropriately 	Indicators: <ul style="list-style-type: none"> □ Clear description and analyses of raw data, themes, and/or codes □ Results are coded and presented in a manner that is clear 	Indicators: <ul style="list-style-type: none"> □ Discussion of results clearly address the research question(s) □ Findings are compared to appropriate prior research □ Limitations of the study are clearly identified □ Recommendations for future studies are clear and detailed □ Inclusion and exclusion of reported data were thoroughly addressed □ Generalizability addresses target population and other possible issues □ Researcher clearly identifies personal perspectives and positions as it relates to the study 	Indicators: <ul style="list-style-type: none"> □ Contains a complete listing of references pertinent to the study □ Appendices are provided
	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 2 (Moderate)	<p>Indicators:</p> <ul style="list-style-type: none"> □ Research question(s) is/are present but not clear □ Purpose statement does not clearly align with research question(s) □ Limited application to theory and/or a conceptual model □ Significance and need for the study is not clearly demonstrated □ Some solutions to challenges are provided □ Literature lacks direct connection to the purpose and justification of the study <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> □ Research design somewhat aligns with the research question(s) □ Limited measures are used to control for participant and researcher bias □ Participants selected vaguely reflect the intended purpose of the study □ Population does not provide a full description of inclusion and exclusion criteria □ Participant criteria are described but lacks detail for replication □ Participant confidentiality is present but not clear □ Data collection is limited or not clear if it was conducted throughout the course of the treatment; preliminary data level is weak □ Sampling technique(s) is/are described but is not replicable □ Researcher(s) provides adequate information for replication of study and setting is adequately described □ Treatment levels are adequately explained and documented □ Not all interview questions are within the purpose of the study or answer the research question □ Measures are mentioned to explain the accuracy of the description of the treatment of collected data □ If appropriate some documents/artifacts are meaningful, accurately described, and cited appropriately <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> □ Incomplete description and analysis of raw data, themes and/or codes □ Results are coded and presented but are not clear <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> □ Discussion of results does not clearly address the research question(s) □ Findings are present but not all are connected to appropriate prior research □ Limitations are mentioned but not clearly presented □ Recommendations are mentioned but not addressed in detail □ Inclusion and exclusion of reported data were addressed but are not thorough □ Generalizability addresses target population and other possible issues but is limited □ Researcher acknowledges and identifies personal perspectives and positions but they are limited or do not relate to the study <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> □ Partial inclusion of references pertinent to the study □ Appendices are incomplete <p>Indicator evaluation</p>

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 3 (Weak)	<p>Indicators:</p> <ul style="list-style-type: none"> □ Research question(s) are not present □ Purpose of the study does not align with research question(s) □ No application to theory and/or conceptual model □ Significance and need for the study is not identified or is opinion based □ Solutions to challenges are not provided □ Literature has no clear connection to the purpose and justification of the study 	<p>Indicators:</p> <ul style="list-style-type: none"> □ Research design does not align with the research question(s) □ No apparent measures are taken to control for participant bias □ Participants selected do not reflect the intended purpose of the study □ Population does not provide a description of inclusion and exclusion criteria □ Participant criteria description is not replicable □ Participant confidentiality is not addressed □ No evidence of data collection was conducted throughout the course of the treatment; no preliminary data is determined □ Sampling technique(s) is/are not provided □ Researcher(s) provide insufficient information for replication of study and setting is not addressed □ Treatment levels are inadequately explained and/or documented □ Interview questions are not within the purpose of the study or do not answer the research question □ Measures are not evident to ensure the accuracy of the description of the treatment of collected data □ If appropriate documents/artifacts do not appear to be meaningful, accurately described, or cited appropriately 	<p>Indicators:</p> <ul style="list-style-type: none"> □ No clear description and analyses of raw data, themes and/or codes □ Results are not coded are not clear 	<p>Indicators:</p> <ul style="list-style-type: none"> □ Discussion of results does not address research question(s) □ Findings are not connected to appropriate prior research □ Limitations are not identified □ Limited or no recommendations to future research is presented □ Inclusion and exclusion of reported data were not addressed □ Generalizability does not addresses target population and other possible issues □ Researcher(s) does/do not identify personal perspectives and positions as it relates to the study 	<p>Indicators:</p> <ul style="list-style-type: none"> □ Does not include references pertinent to the study □ Appendices are not provided <p>Indicator evaluation</p>

APAT: Level of Recommendation (Part II)



Used and modified from Newman, M., Weyant, R., & Hujoel, P. (2007). JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews. *Journal of Evidence-Based Dental Practice*, 7, 147-150. doi:10.1016/j.jebdp.2007.09.014

Appendix I

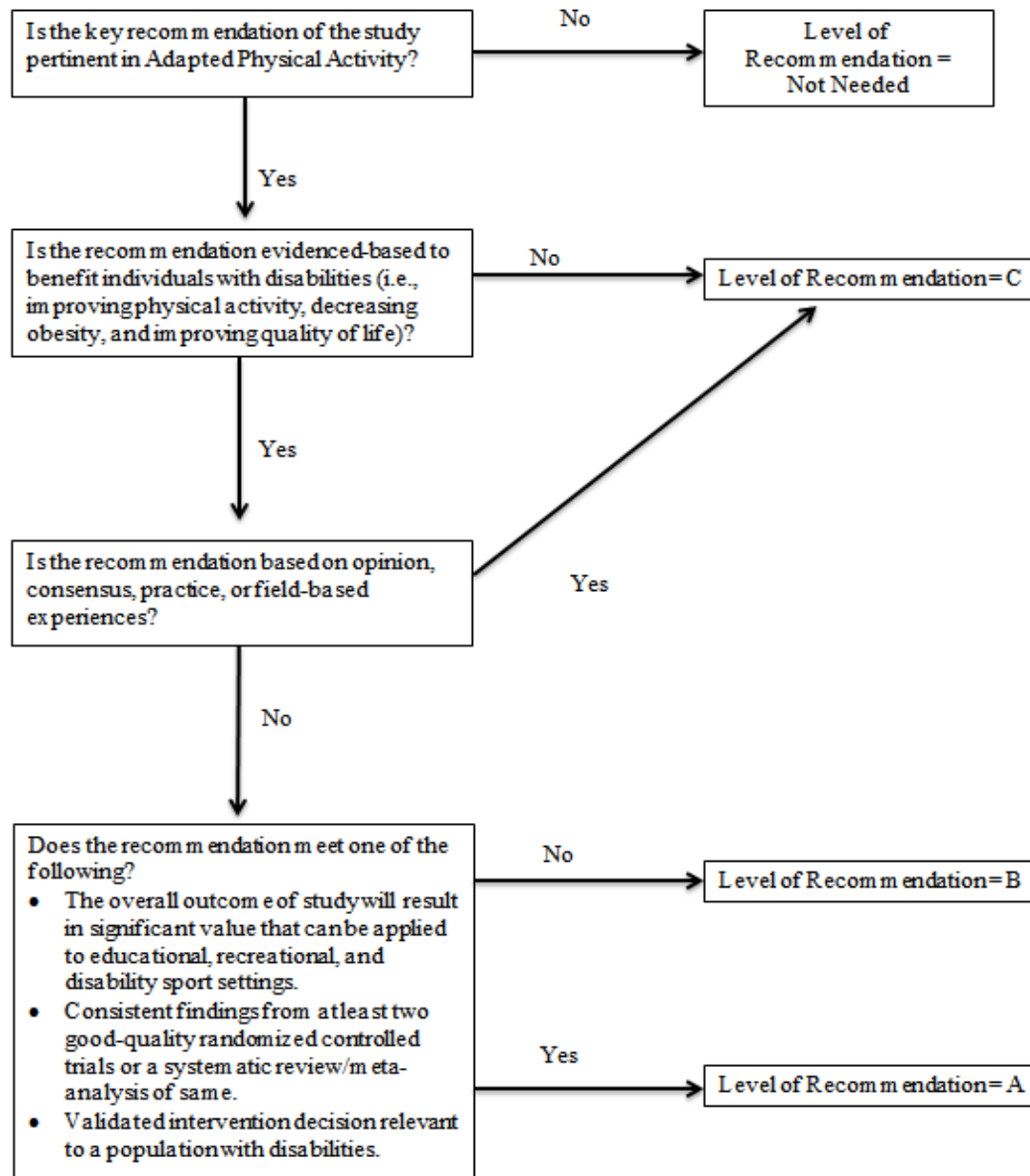
Final Adapted Physical Activity Taxonomy: Single-Subject Design

Part I: Quality of the Study		APAT: <u>Single Subject Design</u> Based on the <u>Strength of Recommendation Format</u>			
Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 1 (Strong)	Indicators: <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are clearly stated Purpose statement for the study clearly aligns with hypothesis(es)/ research question(s) Clear application to a theory or conceptual model Significance and need for the study is clearly demonstrated Solutions to the challenges are provided Literature supports the purpose and justification of the study 	Indicators: <ul style="list-style-type: none"> Research design appropriately aligns with the hypothesis(es)/research question(s) Participants selected clearly reflect the intended study Population is adequately represented Population includes full description of inclusion and/or exclusion criteria Participant criteria are fully described to the extent that is replicable Participant confidentiality is clearly addressed Data collection was conducted throughout the course of the treatment to substantiate trustworthiness; appropriate baseline measurements are obtained Researcher(s) provides sufficient information for replication of the study and includes a clear description of setting Sample technique(s) is/are described in depth Dependent variable (DV) and Independent variable (IV) levels, intervention, and/or conditions are thoroughly explained, documented, and/or replicable DV is quantifiable IV is manipulated under the control of the experimenter Baseline data were systematic and provide evidence for repeated measurement over time Threats to internal validity are thoroughly addressed 	Indicators: <ul style="list-style-type: none"> Percent agreement between observers is $\geq 80\%$, or coefficient r is ≥ 0.7 Analyses of raw data were clearly described A pattern of experimental control is clearly discussed At least 2 different experimental effects over 2 different periods of time are clearly presented 	Indicators: <ul style="list-style-type: none"> Discussion of results clearly addresses the hypothesis(es)/research question(s) Findings are compared to appropriate prior research Limitations of the study are clearly identified Recommendation for future studies are clear and detailed Inclusion and exclusion of reported data were thoroughly addressed Generalizability addresses target population and other possible issues DV demonstrates social importance DV is supported by a magnitude of change IV is practical and cost effective 	Indicators: <ul style="list-style-type: none"> Contains a complete listing of references pertinent to the study Appendices are provided
	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation	Indicator evaluation

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 2 (Moderate)	<p>Indicators:</p> <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are present but not clear Purpose of the study does not clearly align with hypothesis(es)/ research question(s) Limited application to theory and/or conceptual model Significance and need for the study is not clearly demonstrated Some solutions to challenges are provided Literature lacks direct connection to the purpose and justification of the study <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Research design somewhat aligns with the hypothesis(es)/research question(s) Participants selected vaguely reflect the intended purpose of the study Population is underrepresented Population does not provide a full description of inclusion and/or exclusion criteria Participant criteria are described but lacks detail for replication Participant confidentiality is present but not clear Data collection is limited or not clear if it was conducted throughout the course of the treatment; baseline level is weak Sample technique(s) is/are described but is not replicable Researcher(s) provide limited information for replication of study of study and setting is adequately described DV and IV levels, intervention, and/or conditions are adequately explained, documented, and/or replicable DV is quantifiable but subjective IV is partially manipulated under the control of the experimenter Baseline data were described but may not provide evidence for repeated measurement over time Threats to internal validity are minimally addressed <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Percent agreement between observers is between 70% to 79%, or coefficient r is between 0.5 to 0.69 Analyses of raw data were presented but are not clear A pattern of experimental control is adequately discussed Two different experimental effects over 2 different periods of time are presented <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Discussion of results does not clearly address the hypothesis(es)/research question(s) Findings are present but not all are connected to appropriate prior research Limitations are mentioned but not clearly presented Recommendation are mentioned but not addressed in detail Inclusion and exclusion of reported data were addressed but are not thorough Generalizability addresses target population and other possible issues but is limited DV demonstrates limited social importance DV's magnitude of change is provided but is limited IV may be practical but is not cost effective <p>Indicator evaluation</p>	<p>Indicators:</p> <ul style="list-style-type: none"> Partial inclusion of references pertinent to the study Appendices are incomplete <p>Indicator evaluation</p>

Strength of Quality	Domain I: Introductory Section	Domain II: Method Section	Domain III: Results Section	Domain IV: Discussion Section	Domain V: Other
Level 3 (Weak)	<p>Indicators:</p> <ul style="list-style-type: none"> Hypothesis(es)/ research question(s) is/are not present Purpose of the study does not align with hypothesis(es)/ research question(s) No application to theory and/or conceptual model Significance and need for the study is not identified or is opinion based Solutions to challenges are not provided Literature has no clear connection to the purpose and justification of the study 	<p>Indicators:</p> <ul style="list-style-type: none"> Research design does not align with the hypothesis(es)/research question(s) Participants selected do not reflect the intended purpose of the study Population does not provide a description of inclusion and/or exclusion criteria Participant criteria description is not replicable Participant confidentiality is not addressed Sample technique(s) is/are not provided No evidence of data collection was conducted throughout the course of the treatment; no baseline level is determined Researcher(s) provide no information for replication of study DV and IV level, intervention, and/or conditions are inadequately explained, documented, and/or not replicable DV is not quantifiable IV is not manipulated under the control of the experimenter or is not addressed Baseline data were not addressed and there is no evidence of repeated measurement over time Threats to internal validity are not addressed 	<p>Indicators:</p> <ul style="list-style-type: none"> Percent agreement between observers is $\leq 69\%$, or coefficient r is ≤ 0.49 Analyses of raw data were missing A pattern of experimental control is not discussed Only 1 different experimental effect over 1 period of time is presented 	<p>Indicators:</p> <ul style="list-style-type: none"> Discussion does not address the hypothesis(es)/research question(s) Findings are not connected to appropriate prior research Limitations are not identified Limited or no recommendations to future research is presented Inclusion and exclusion of reported data were not addressed Generalizability does not addresses target population and other possible issues DV does not demonstrate social importance DV does not report a magnitude of change IV is neither practical or cost effective 	<p>Indicators:</p> <ul style="list-style-type: none"> Does not include references pertinent to the study Appendices are not provided <p>Indicator evaluation</p>

APAT: Level of Recommendation (Part II)



Used and modified from Newman, M., Weyant, R., & Hujoel, P. (2007). JEBDP Improves grading system and adopts strength of recommendation taxonomy grading (SORT) for guidelines and systematic reviews. *Journal of Evidence-Based Dental Practice*, 7, 147-150. doi:10.1016/j.jebdp.2007.09.014