2016

Effects of exercise-based interventions for children with Autism Spectrum Disorder (ASD) :a systematic review and meta-analysis

Sam Ka Lam Sam
Hong Kong Baptist University

HONG KONG BAPTIST UNIVERSITY

Master of Philosophy

THESIS ACCEPTANCE

DATE: December 23, 2015

STUDENT'S NAME: SAM Ka Lam Sam

THESIS TITLE: Effects of Exercise-based Interventions for Children with Autism Spectrum Disorder (ASD): A Systematic Review and Meta-Analysis

This is to certify that the above student's thesis has been examined by the following panel members and has received full approval for acceptance in partial fulfillment of the requirements for the degree of Master of Philosophy.

Chairman: Prof. Zhou Qiming
Professor, Department of Geography, HKBU
(Designated by the Dean of Social Sciences)

Internal Members: Prof. Chung Pak-Kwong
Professor, Department of Physical Education, HKBU

Prof. Chow Bik Chu
Professor, Department of Physical Education, HKBU

External Member: Prof. Sit Hui Ping Cindy
Associate Professor
Department of Sports Science and Physical Education
The Chinese University of Hong Kong

Issued by Graduate School, HKBU

Effects of exercise-based interventions for children with Autism Spectrum Disorder (ASD): A Systematic Review and Meta-Analysis

SAM Ka Lam, Sam

A thesis submitted in partial fulfilment of the requirements

for the degree of

Master of Philosophy

Principal Supervisor: Prof. CHOW Bik Chu

Hong Kong Baptist University

December 2015

DECLARATION

I hereby declare that this thesis represents my own work which has been done after registration for the degree of MPhil at Hong Kong Baptist University, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

Signature: ______________________

Date: December 2015

ABSTRACT

Autism Spectrum Disorder (ASD) is a prevalent neurological syndrome observed even in childhood stages. Children with autism have a certain range of deficits on social interaction, communication and coordination. Exercise program has been commonly adopted as an intervention for ASD children on developing both of their motor and social skills, as well as maintaining a certain fitness level. This project is a systematic review and meta-analysis study which summarizes the recent fifteen years (2000-2015) of control-trial exercise-based interventions for ASD children and evaluates their effectiveness from three aspects (according to the outcome measures): 1) Exercise Performance & Sport/Skill-related Fitness (EXE); 2) Physiological & Biometric Indicator (PHY); and 3) Social Cognition & Psychological Well-being (SOC). The results showed a positive effect in all three aspects: .763, .412, and .505 respectively. Further studies were recommended on investigating the mechanism explaining the psycho-social and physiological effects of exercise programs on ASD children.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors:

Prof. Chow Bik Chu (Principal Supervisor)

Dr. Tong Kwok Keung, Tom (Co-supervisor)

Prof. Pan Chien Yu (Co-supervisor)

Special thanks to:

Prof. Lo Sing Kai (Epidemiology)

Prof. Lam Tak C., Eddie (Measurement)

Dr. Andrew W. Smith (Rehabilitation Science)

Dr. Ho Fuk Chuen (Special Education)

Ms. Lam Sze Ching, Cici (Educational Psychologist)

Dr. Tsang Kwan Lan, Vicky (Occupational Therapist)

Ms. Chan Yee Kwan, Florence (Speech Therapist)

Mr. Chau Chung Hei (Speech Therapist)

Prof. Lau Wing Chung, Patrick (Physical Education)

Prof. Cheung Siu Yin (Physical Education)

Prof. Chung Pak Kwong (Physical Education)

Without your support, this project would not be accomplished; thanks all.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBLOS

α	Alpha
β	Beta
d	Cohen’s d
p	probability; p value
Q	Cochran Q
df	Degree of Freedom
CI	Confidence Interval

LIST OF ABBREVIATIONS

ABC	Aberrant Behavior Checklist
AHCI	Arts and Humanities Citation Index
APA	American Psychological Association
ASD	Autism Spectrum Disorder
ATEC	Autism Treatment Evaluation Checklist
BIA	Bioelectrical Impedance Analysis
BMI	Body Mass Index
BOTMP	Bruininks-Oseretsky Test of Motor Proficiency
CARS	Childhood Autism Rating Scale
CONSORT	Consolidated Standards for Reporting of Trials
CUHK	Chinese University of Hong Kong
DSM	Diagnostic and Statistical Manual of Mental Disorders
EXE	Exercise Performance & Sport/Skill-related Fitness
FTS	Floor to Stand
GARS	Gilliam Autism Rating Scale
HAAR	Humphries Assessment of Aquatic Readiness
HKU	University of Hong Kong
HKUST	Hong Kong University of Science and Technology

ID	Intellectual Disability
IPU	Isometric Push-Up
MIT	Massachusetts Institute of Technology
M-PEDI	Multidimensional Pediatric Evaluation of Disability Inventory
NLM	National Library of Medicine
OU	Open University
PACER	Progressive Aerobic Cardiovascular Endurance Run
PHY	Physiological & Biometric Indicator
PolyU	Hong Kong Polytechnic University
SCI	Science Citation Index
SIS	Sensory Information Storage
SnR	Sit and Reach
SOC	Social Cognition & Psychological Well-being
SP	Sensory Profile
SRS	Social Responsiveness Scale
SSBS	School Social Behavior Scale
SSCI	Social Sciences Citation Index
TPCIS	Timberlawn Parent-Child Interaction Scale
TSIF	Test of Sensory Integration Function

CHAPTER 1

Introduction

Autism Spectrum Disorder (ASD), according to the latest Diagnostic and Statistical Manual of Mental Disorder, 5th Edition, (DSM-V) (APA, 2013), is an individual encountering deficiencies in social-emotional reciprocity, verbal and/or non-verbal communication, and developing, maintaining and understanding relationships. The overall prevalence of ASD is around 4.0% worldwide, but only 0.3 to 1.0% were diagnosed, which covers a certain range of neuropsychological conditions that affect both individual and group functioning (APA, 2013). ASD is a behaviorally defined syndrome with a wide variety of both genetic and non-genetic causes (Muhle, Trentacoste, & Rapin, 2004). There is no current evidence that ASD is linked exclusively to any particular genetic or non-genetic cause (Muhle et al., 2004). Matson and Kozlowski (2011) suggested that the increasing prevalence of ASD may be related to environmental and cultural factors, as well as improvements and variability in diagnostic techniques (Matson & Kozlowski, 2011). King and Bearman (2009) found that approximately 25% of the rise in ASD over the past two decades can be attributed to diagnostic accretion, i.e., some children who would have been diagnosed with Intellectual Disability (ID) previously are now given a diagnosis of both ID and ASD (King & Bearman, 2009). Nevertheless, the purported growing awareness of ASD has been attributed to 15% of the rise in prevalence while geographic associations account for 4% of the rise (King & Bearman, 2011). Another 10% of the increase may relate to people having children when they are older (King, Fountain, Dakhlallah, & Bearman, 2009). Some research has found that children born to parents older than 35 have a higher risk of being diagnosed with autism, with the mother's age being a more critical factor (King et al., 2009). Despite these explanations, King and Bearman (2011) reported that 46% of the increase in ASD cannot be explained (King & Bearman, 2011).

According to Gentile's study (2013), ASD is one of the most common pediatric conditions in the United States, with prevalence rates of one in 88 for both sexes and one in 54 for boys (Gentile et al., 2013), and the prevalence is still rising (Matson & Kozlowski, 2011). Males develop the condition four times as often as females (Belfer, 2008; Matson & Kozlowski, 2011), and each year, approximately 36,500 new cases emerge, amounting to a national US total of cases (Centers for Disease Control, 2009), costs for the family and society at large can be astronomical. Ganz (2007) estimated that the societal cost for and with autistic disorder across his or her lifetime was 3.2 million USD. The majority of this cost is accounted for by adult care and a loss of productivity for the individuals with autism, as well as their parents (Ganz, 2007; Gentile et al., 2013).

Intellectual Disability (ID) commonly occurs with ASD, with an estimated 25.8% of children diagnosed with ASD also having a diagnosis of ID (Chakrabarti & Fombonne, 2001; Chakrabarti & Fombonne, 2005). Moreover, Chakrabarti and Fombonne (2001; 2005) showed in their study that 70% of children who met the strict criteria for autism disorder had also given a diagnosis of ID (Chakrabarti & Fombonne, 2001; Suniti Chakrabarti & Fombonne, 2005). Furthermore, the severity of ID is positively correlated with the severity of autistic symptoms (Bouras, Holt, Day, & Dosen, 1999; de Bildt et al., 2004; La Malfa, Lassi, Bertelli, Salvini, & Placidi, 2004). However, individuals with ASD show greater difficulties with social skills compared to individuals without the condition but with the same level of intellectual functioning (Matson et al., 2009).

Regardless the prevalence rate and the cognitive issue, although a clear definition was given in DSM-V on ASD, according to recent studies (Silver & Rapin, 2012; Srinivasan, Pescatello, & Bhat, 2014) identified, in the field of exercise science, children who diagnosed with ASD display deficiencies preliminary on two major areas of practical functioning, which are 1) socialization and 2) motor performance. In the perspective of child development, those are highly related to the developmental levels on social and motor skills.

Social skill of children with ASD

The Theory of Mind (Baron-Cohen et al., 1985) is used to describe the inability of individuals with autism to understand the perspectives, beliefs, feelings and impressions of others. The Theory of Mind also illustrates how children with autism have difficulty making the connection between mental states and ensuing behaviors (Baron-Cohen et al., 1985). For example, children with autism often find it hard to understand complex social emotions such as guilt, pride, embarrassment, empathy, loneliness and surprise and do not usually understand implicit social norms and rules (Bauminger, 2002). Many children with autism seldom initiate or maintain social interactions and can have severe deficits in the areas of imitation, play skills and responsiveness to others (Maione & Mirenda, 2006). They often have impaired eye- gaze, difficulty with joint-attention, very few verbal initiations and an inability to develop age appropriate friendships (Dawson et al., 2004). Fortunately for those with autism, social interest expands and social skills continue to develop during adolescence and adulthood (McGovern & Sigman, 2005).

There is often a relationship between problem behaviors and poor social development (Licciardello, Harchik, & Luiselli, 2008). The stereotypic behaviors common among those with autism frequently alienate peers due to the ‘highly unusual and stigmatized nature’ of such behaviors (Lee, Odom, & Loftin, 2007). This in turn impedes the child’s ability to develop proficient social skills due to limited exposure to social situations (Lee et al., 2007). Even when children with ASD are in close proximity to their typical peers in the schoolyard, they will not usually join in their peers’ activities (Owen-DeSchryver, Carr, Cale, & Blakeley-Smith, 2008) and generally prefer to engage in solitary activities (Anderson, Moore, Godfrey, & Fletcher-Flinn, 2004). Intervention is habitually necessary to acquire functional social skills. Social-skills training is a common educational objective for school-aged children with ASD (Licciardello et al., 2008).

Young individuals with ASD often have aberrant social behavior and excessive desire for sameness (Hoeksma, Kemner, Verbaten, & van Engeland, 2004) as well as a limited range of language skills that hinder participation in social interactions (Thiemann & Goldstein, 2004). For children with ASD, an inability to consider the perspectives of other people is common and the child is often perceived as precocious or immature (Gillberg, 2002). ASD children commonly display a near obsession with truth telling and this habit often places the child in socially uncomfortable situations. The child with ASD does not usually have the ability to differentiate between what is appropriate to say and what is not. Contextually appropriate body language and facial expressions are also a common problem in youth with ASD. It is not uncommon for a person with ASD to appear ‘stone-faced’ or depressed during an intense conversation and then laugh or smile once the conversation is over. Standing beside, behind or in unusual proximity to another person throughout a conversation is common (Gillberg, 2002). Children with ASD typically do not have many friends, despite an increased interest in forming relationships in the adolescent years (Gillberg, 2002).

Exercise and motor performance of children with ASD

It has been suggested that motor impairments are widespread within ASD population, and therefore can justifiably be considered a symptom of ASD (Bhat, Landa, & Galloway, 2011; Fournier, Hass, Naik, Lodha, & Cauraugh, 2010). The range of symptoms that people with ASD may display necessitates that treatments address multiple deficiencies simultaneously, while being tailored to individual needs, which makes knowledge of ASD-specific interventions that have an overall positive impact indispensable. Standard ASD treatment, such as behavioral intervention program, is typically aimed at stimulating cognition, language and social development while trying to suppress or eliminate maladaptive behavioral patterns such as rigidity and stereotypical movements (Sowa & Meulenbroek, 2012; Tanaka et al., 2010). Some methods widely used in behavioral intervention programs focus on developing communication, social and cognitive skills to treat children with ASD. However, recent research (i.e., Chan, Han, & Cheung, 2014; Tanaka et al., 2010) suggests that some alternative therapeutic choices that include sports, exercise and other physical activities can be of help to the autistic child. Individuals with autism need to improve their physical fitness and motor coordination; they need programs that can motivate them to participate in games. Since autistic children generally demonstrate poor motor skills, children and adolescents with autism spectrum and high functioning autism have delays or disorders in overall motor development, including locomotors and object control (Berkeley, Zittel, Pitney, & Nichols, 2001). Consequently, rehabilitation programs should emphasize fundamental motor skills and patterns of movement, individual games and sports and developmental activities that increase physical proficiency. Some research sectors (i.e., Lord & McGee, 2001) indicated that the motor functioning of individuals with ASD has been a neglected area. On the other hand, the previous studies indicated that participation in physical activity has been shown to have multiple benefits, including reduction of stereotypic behavior. Physical exercise offers a variety of benefits for individuals with autism. It has been found to help decrease some of the many challenges (i.e., maladaptive behaviors, coordination difficulties, and etc.) those children with autism face (Berkeley et al., 2001; O Connor, French, & Henderson, 2000).

According to Baranek (2002), developmental motor delays are present from infancy in those with autism and become more apparent with age. In one of her studies it was concluded that babies later diagnosed with autism were significantly delayed in the development of gross motor milestones at six months and thereafter throughout the first year of life. This is serious, since a lag in motor development during the foundational years can impact the subsequent development of other such skills (Baranek, 2002).

Motor skills are critical since they serve as foundations for other skills such as perception, incorporation of outside stimuli and sequencing skills (Blakemore-Brown, 2002). According to Vernazza-Martin (2005), motor problems in autism stem from difficulty organizing motor actions toward a goal as well as problems of anticipatory function (Vernazza-Martin et al., 2005). Official terms used to describe motor problems associated with autism are ‘apraxia’, ‘praxis’ and ‘dyspraxia’, all of which indicate the individual’s difficulty with motor planning and perception (Hewetson, 2002). Interestingly, children with autism do not always develop hand specialization, i.e., they do not have one dominant hand to use in motor tasks. This may play a role in social imitation difficulties as well as goal directed motor tasks, such as reaching and grasping (Hewetson, 2002).

Kanner described the children with autism he observed as being ‘clumsy in gait and gross motor performance’ (Reid & Collier, 2002). Likewise, children with autism often experience challenges when trying to perform coordinated movements of different parts of the body simultaneously (Peeters & Gillberg, 1999). Many children with autism demonstrate low muscle tone, oral-motor difficulties and repetitive motor movements (Baranek, 2002). Their motor skills can appear ‘awkward’ (Nash & Collins, 2006), and motor deficiencies in autism are commonly misinterpreted as ‘general clumsiness’ (Baranek, 2002). They often demonstrate unusual appearance with regard to posture and gait (Klin, Volkmar, Sparrow, Cicchetti, & Rourke, 1995). Fine motor skills as well as gross motor skills show marked deficiency in those with autism and this is often a cause for anxiety when performing motor tasks (Peeters & Gillberg, 1999). Studies have suggested that games involving ball skills are exceptionally difficult for most children with ASD (Reid & Collier, 2002). Reid and Collier (2002) have noted that the movement skills of children with ASD are typically poorly developed or delayed (Reid & Collier, 2002).

Much of the research conducted on brain structure in autism has focused on the dysfunction of the cerebellum (Fatemi, Merz, & Realmuto, 2003). One of the cerebellum's key functions is to regulate motor functioning and to correlate motor acts with emotional and motivational processes in the brain (Bauman & Kemper, 2005). According to Bauman and Kemper (2005), an individual with cerebellum dysfunction have impaired motor movements. In keeping with their theory, emotions and motivations within the individual's thought process will be unable to 'connect' with motor acts. In addition, they explain that possible cerebellum abnormalities in those with autism could cause an overshoot or inability of the motor system to perceive parameters of movement. Considerable literature has described the motor delays and challenges associated with the specific diagnosis of ASD. For children with AS, delays in motor development during the first year are common. These motor delays include late onset or absence of crawling, sitting unsupported, rising to a standing position and delayed onset of walking. General clumsiness and poorly coordinated gross motor movements are almost always present in children with ASD at any age (Gillberg, 2002). As children get older, other motor delays often become apparent, such as poor balance and general lack of motor coordination. Motor skill deficiencies in ASD can affect the child's ability to perform school related tasks such as writing, art, social skills and vocational skills. Children with ASD may also have difficulty being involved in games with motor demands, keeping them from participating with their peers (Rotheram-Fuller, Kasari, Chamberlain, & Locke, 2010). Up until the age of ten, it may be extremely difficult for a child with AS to engage in common motor tasks, such as bicycle riding, swimming or catching and kicking a ball (Gillberg, 2002).

Current systematic review and meta-analysis studies on exercise intervention of children with ASD

A literature search was performed to identify the recent systematic review and meta-analysis on the effect of non-clinical interventions in children with ASD. The keywords used were “autism”, “autism spectrum disorder”, “ASD”, “motor function”, “motor skill”, “motor performance”, “exercise”, “fitness”, “physical activity”, “therapy”, “treatment”, and “intervention”, pairing with the study terms - “systematic review” or “meta-analysis”. The electronic databases included MEDLINE, PubMed, EMBASE, CINAHL, PsychINFO, EBSCOhost, and the Cochrane library. The search for the studies included the years from 2005 to 2015.

After an initial search for the studies of systematic review and meta-analysis, fifteen were identified. Two of them were guidelines, and were excluded because of the incompleteness as a study. Three of them did not meet the definition of systematic review and meta-analysis in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*. Eleven studies were finally included for investigation (Figure 1).

**PRISMA is an evidence-based 27-item checklist for reporting in systematic reviews and meta-analyses (www.prisma-statement.org).*

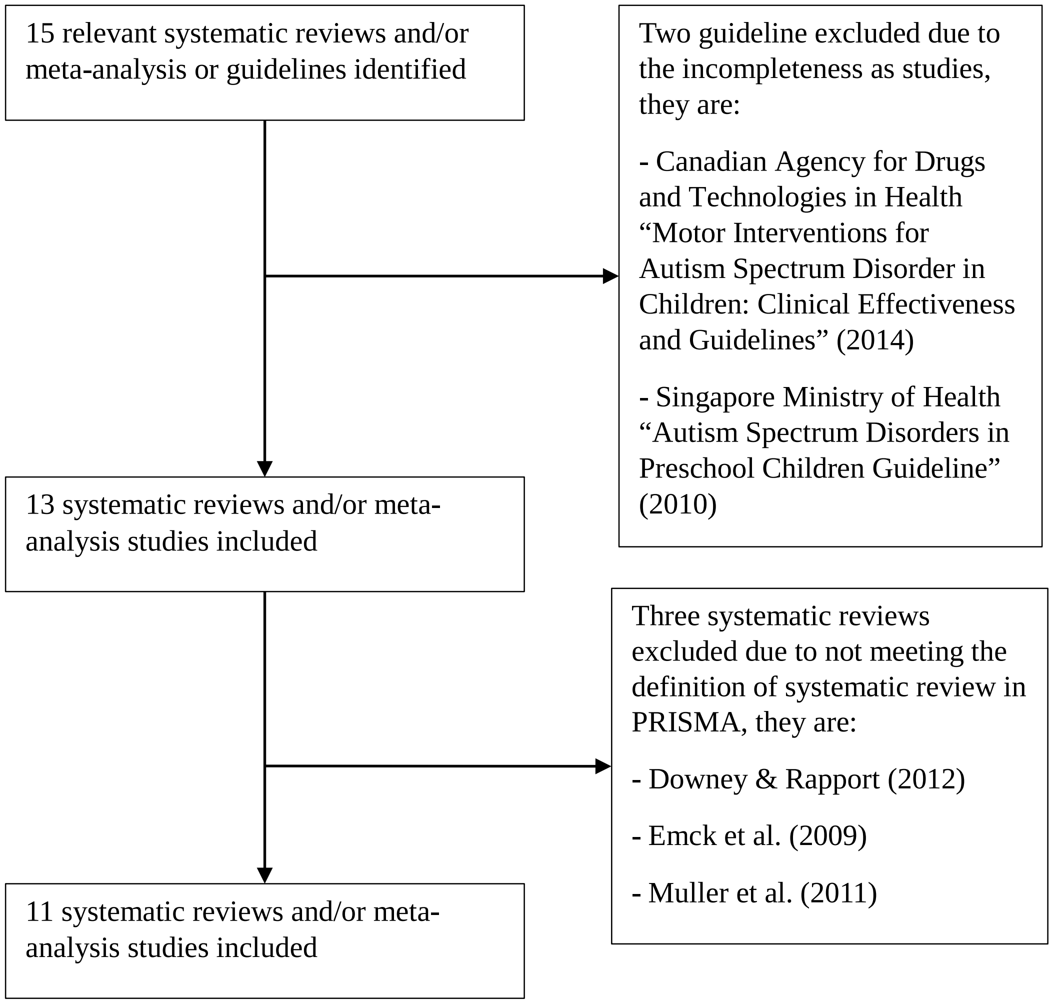


Figure 1. Flow chart of study inclusion

The quality assessment was performed using a particular scale for systematic review and meta-analysis study - A Measurement Tool to Assess Systematic Review (AMSTAR)*. The rating procedure was conducted by two independent reviewers in the field of education and psychology. The results of the quality rating was shown in Table 1.

Note. **AMSTAR is 11-item checklist for assessing the quality of systematic reviews (Amstar.ca).*

Table 1. Critical appraisal of study quality - AMSTAR

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Fournier et al. (2010)	1	1	1	1	1	1	1	0	0	1	0
Lang et al. (2010)	1	1	1	0	0	1	1	1	0	0	0
Leonard et al. (2014)	1	0	1	0	0	1	1	0	0	0	0
Miyahara et al. (2013)	1	0	1	0	0	1	1	1	0	0	0
Mortimer et al. (2014)	1	1	1	0	0	1	1	1	0	0	0
Nickl-Jockschat et al. (2012)	1	0	0	0	0	1	0	0	1	0	0
Petrus et al. (2008)	1	1	1	1	0	1	1	1	0	0	0
Philip et al. (2012)	1	0	1	1	0	1	0	0	1	0	0
Sowa et al. (2012)	1	0	0	0	0	1	0	0	0	1	0
Stanfield et al. (2008)	1	1	1	1	0	1	1	0	1	1	0
Virues-Ortega et al. (2013)	1	0	1	1	0	0	1	1	1	1	0

Q1: Prior design provided (0/1)
Q2: Duplication and data extraction (0/1)
Q3: Search literature comprehensively (0/1)
Q4: Publication as inclusion criterion (0/1)
Q5: Study list complied (0/1)
Q6: Study characteristics provided (0/1)
Q7: Quality assessment (0/1)
Q8: Proper use of quality assessment (0/1)
Q9: Proper use of combination of findings (0/1)
Q10: Publication bias assessed (0/1)
Q11: Conflict of interest included (0/1)

Among the eleven included studies, five of them were cross-sectional comparison studies: two of them were compared the motor function between ASD and non-ASD individuals; two of them compared the structural change; one of them compared the brain activation. Five of them were longitudinal cases series (including case studies, control trials): three of them were tested the effect of exercise; one of them tested the effect of hydrotherapy; one of them tested the effect of TEACCH program (TEACCH stands for Treatment and Education of Autistic and Related Communication Handicapped Children, TEACCH is a clinical, training, and research program based at the University of North

Carolina Chapel Hill, it was developed by Drs. Eric Schopler and Robert Reichler in the 1960s; it has become a model for other autism programs around the world nowadays). The remaining one was a meta review, which reviewed 12 systematic reviews from inception to 2012. Detailed characteristics, including age groups and number of studies included, please see the Table 2 below (Only the first author was shown under the field of study).

Table 2. Characteristics and the quality of the included studies

Study	Number of included studies	Design of included studies	Theme	Age group	Quality*
Fournier et al. (2010)	51	Comparison	Comparison of Motor function	Child Adolescence Adult	High (8)
Lang et al. (2010)	18	Case series	Effect of exercise	Child Adolescence Adult	Medium (6)
Leonard et al. (2014)	43	Comparison	Comparison of Motor function	Infant Child Adolescence	Medium (4)
Miyahara et al. (2013)	12	Systematic review	Meta-review	Non-applicable	Medium (5)
Mortimer et al. (2014)	4	Case series	Effect of hydrotherapy	Child Adolescence	Medium (6)
Nickl-Jockschat et al. (2012)	16	Comparison	Comparison of structural change	Adolescence Adult	Low (3)
Petrus et al. (2008)	7	Case series	Effect of exercise	Child Adolescence	Medium (7)
Philip et al. (2012)	3	Comparison	Comparison of brain activation	Child Adolescence Adult	Medium (5)
Sowa et al. (2012)	16	Case series	Effect of exercise	Child Adolescence Adult	Low (3)
Stanfield et al. (2008)	46	Comparison	Comparison of structural change	Child Adolescence Adult	High (8)
Virues-Ortega et al. (2013)	13	Case series	Effect of TEACCH program	Child Adolescence Adult	Medium (7)

Quality: Based on AMSTAR scores [0-3: Low; 4-7: Medium; 8-11: High]

As abovementioned, three case series were evaluated the effect of exercise, including Lang et al. (2010), Petrus et al. (2008), and Sowa and Meulenbroek (2012).

Lang et al. (2010) reviewed eighteen studies on the effect of exercise on behavior, academic, and fitness of 64 individuals (from children to adults; age: 2-41 years) with autism. The sample size of included studies was small, and the majority of studies used a time-series (repeated measures) analysis. The intervention content consisted of serial exercises, such as jogging (majority of the cases), by adopting physical instruction, teaching model, feedback reinforcement, and contingency management. The duration/length of intervention varied widely (8-30 minutes per session; number of sessions were unknown). The quality of this paper was medium (6) according to AMSTAR rating.

Petrus et al. (2008) reviewed seven studies on the effect of exercise on stereotypic behaviors of 25 individuals (from children to adolescences; age: 4-15 years) with autism. The studies included were small in sampling size (n<6), and most of them used case-series or single case study designs. The intervention content consisted of walking, jogging, ball game and hydrotherapy. The duration/length of intervention was 6-20 minutes per session, and the frequency and total number of sessions were unknown. The quality of this paper was medium (7) according to AMSTAR rating.

Sowa and Meulenbroek (2012) reviewed sixteen studies on the effect of exercise on motor and social skills of 133 individuals (from children to adults; age: 4-41 years) with autism. The sample sizes of included studies were small (n<7), for those studies, seven of them examined the individualized exercise programs; nine of them evaluated the group exercise programs. The intervention content consisted of walking, jogging, motor training, weight training, swimming, cycling, aerobic activity, as well as hippotherapy and hydrotherapy. The duration/length of intervention was 20-90 minutes per session; 1-2 times per week; 2-12 weeks in total. The effect sizes of those included studies were calculated and given a positive effect for the overall outcome, specifically, individualized programs exhibited better effects comparing to group programs. The quality of this paper was low (3) according to AMSTAR rating.

Among all three studies, only Sowa and Meulenbroek (2012)'s study involved the meta-analysis procedure, however, control groups were absent while calculating the effect size, therefore, confounding variables, such as maturation, repeated practicing, might be present, and played a crucial role on affecting its interpretation. Other than that, the remaining two studies were just systematic reviews, they provided a valuable evaluation, as well as a comparatively scientific protocol on assessing the quality of individual studies. However, the outcomes (or outcome measures) in those studies could not be quantified, the effects could not be combined numerically, and should not give a clear cut-point for clinical/non-clinical decision making. Hence, the current study, with the objective on evaluating the effect (qualitatively and quantitatively) of control-trial exercise-based interventions of children with autism, deemed vital.

Purpose of study

In order to fill out the knowledge gap (no quantitatively combined effects on evaluating the outcomes of exercise-based type of interventions) described above, a systematic review (Chapter 2) followed by a meta-analysis (Chapter 3) were conducted. The ultimate goal (purpose) of the current study was to evaluate the effectiveness, holistically on both qualitative and quantitative perspectives, of control-trial exercise-based interventions for children with ASD.

CHAPTER 2

Systematic Review

Systematic review is a scientific procedure that can be adopted to summarize and appraise the results of implications. The purpose of this systematic review study was to facilitate evidence-based practice in the area, hence, a systematic review of interventions designed to increase the exercise behavior of children with ASD was conducted initially. The specific aim of this review was to describe the characteristics and exercise interventions of the included studies (e.g., participants, how exercise behaviors were taught, and benefits of increased exercise). A review of this type of studies primarily intends to guide and inform practitioners/educators as they develop educational-based exercise programs for individuals with ASD. A secondary aim is to build upon the existing database so as to stimulate future research efforts aimed at using exercise to improve both physical and psychological health of individuals with ASD.

Methods of the systematic review

Few key stages were involved/undertaken in this systematic review, which included the procedures of literature searching, article screening, data extraction and critical appraisal; the step-by-step procedure was summarized in Figure 2 below.

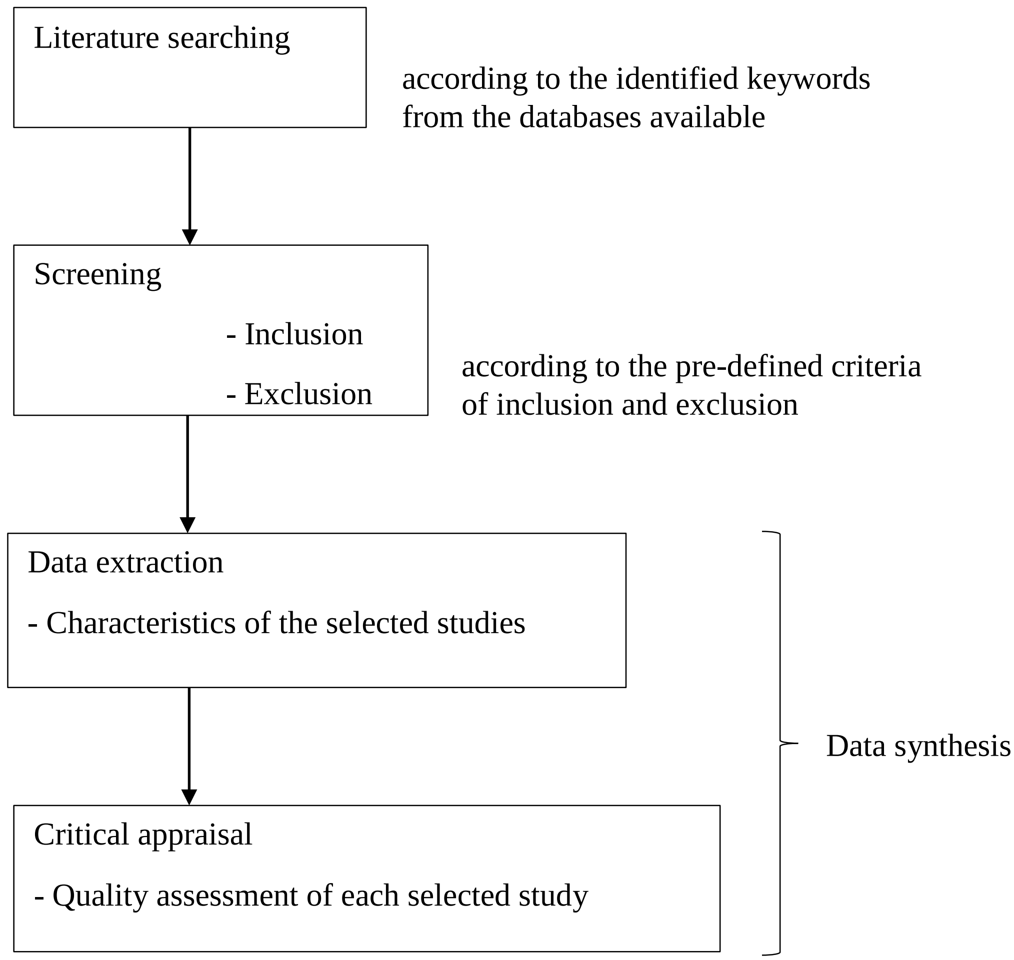


Figure 2. Overview of the systematic review procedure

Search literature

The search covered all electronic databases available on six institutions; 4 in Hong Kong (The Chinese University of Hong Kong, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, University of Hong Kong), 1 in United States (Massachusetts Institute of Technology), and 1 in United Kingdom (Open University). Those databases included MEDLINE, PubMed, EBSCOhost, CINAHL, ProQuest, Scopus, PsycINFO, Cochrane library, Web of Science, and etc. The bibliographies of the initial retrieved studies were search for additional studies, and the "related articles" option was exploited to identify relevant articles on the Web of Science. Google Scholar was also included for further investigation. Keywords for the search were "autism", "autism spectrum disorder", "ASD", matching with "exercise", "fitness", "physical activity", "aerobic", "motor function", "motor skill", "motor performance", or/and "intervention" or "treatment". The search was carried out from 1 January 2015 to 30 April 2015: 1) initial search (January 2015); 2) first phase screening – title and abstract (February 2015); 3) second phase screening – full-text reading (February – April 2015); and 4) confirmation of the list of selected studies and data extraction (March – April 2015).

Inclusion criteria

To be included in this systematic review, studies were required to meet with all the following selection criteria: 1) experimental design published in peer-reviewed journals or non-peer reviewed journal article reporting an empirical intervention; 2) included children with an ASD diagnosis (APA, 2013); 3) the interventions described had to involve exercise or some kind of physical activity; and 4) outcome measures were examined by validated instruments.

Exclusion criteria

Studies were excluded from the systematic review if: 1) non-control-trial studies were applied; 2) articles were written in non-English; 3) those studies were unpublished, i.e., conference papers, manuscript drafts; and 4) studies were published before the year of 2000.

Data extraction

Data were extracted systematically by a standardized data extraction form to minimize bias throughout the data extraction process (Appendix 1). The design of the data extraction form was based on criteria recommended in the National Health Service Centre for Reviews and Dissemination (CRD): CRD’s Guidance for Undertaking Reviews in Health Care (2009) and Report Undertaking Systematic Reviews of Research on Effectiveness (2001). Data extraction included: 1) Study details; 2) Study population; 3) Details of the intervention; 4) Outcome measures; 5) Analysis; and 6) Results.

Quality assessment

Quality assessment aims to assess the study quality of the included studies that is an important part of the systematic review process (Moher et al., 1995). As the study quality might directly influence the result of systematic review, the risk of bias would be enlarged if studies of poor quality were included. Critical appraisal offers a systematic way of assessing the validity, results and usefulness of studies (Hill & Spittlehouse, 2001).

In this application, the quality of primary studies was assessed by the Jadad Score (Jadad et al., 1996). This covered study design, randomization method, method of blinding and description of dropouts, etc. (Appendix 2) Researchers confirmed that the scale was easy to use and time-saving since it just took less than five minutes to score a trial study (Jadad et al., 1998). The scale provided consistent measurements and had construct validity (Jadad et al., 1998). The score may range from 0 to 5 with the worst quality as 0 and the best quality as 5. A trial that was classified as adequately for reporting should score at least three of five points, this cut-off point was recommended by the authors of the scale (Jadad et al., 1996).

However, the Jadad Score (Jadad et al., 1996) focused on the internal validity of the trial such as concealment of allocation. Another appraisal tool, modified Consolidated Standards for Reporting of Trials (CONSORT), adapted by Schulz, Altman and Moher (2010) was also employed (Appendix 3) since this tool could also focus on the external validity of the trial such as description of primary objective, program dosage and variables (Schulz, Altman, Moher, & Group, 2010). The studies that could not maintain the standard would be excluded by critique of the primary studies via questioning. According to the appraisal tool of modified CONSORT, ten questions were used to assess the quality of the trial, which covered study design, random assignment, allocation concealment, groups similar at baseline, eligibility criteria specified, sample size, outcomes, interventions, statistical methods, and program specific features (Schulz et al., 2010).

Data synthesis

In this phase, results of included primary studies were collated and summarized. A summary of the effects of interventions was generated to assist in investigating whether exercise effects on ASD children were consistent across the included studies, and the reasons for any apparent differences. The data synthesis allowed the researchers to generate meaningful conclusions from the reviews. Descriptive data synthesis would be adopted, and the following characteristics were included: 1) author of study; 2) study type; 3) diagnostic criteria; 4) inclusion criteria; 5) mean(range) age; 6) sample size; 7) percentage of female participants; 8) type of interventions; 9) comparators; 10) duration of study; 11) outcome measures; and 12) results. These factors would be summarized clearly in tabulation form National Health Services Centre for Reviews and Dissemination 2009 (Moore et al., 2009).

Results of the systematic review

Identification of primary studies

The search of databases available on 6 institutions: The Chinese University of Hong Kong, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, University of Hong Kong, Massachusetts Institute of Technology and Open

University, were performed. The search results are summarized in Table 3.

Table 3. Number of studies identified at search of institutional subscribed databases

	Institution					
Keyword	CUHK	PolyU	HKUST	HKU	MIT	OU
autism	161	189	56	206	105	168

Among all 6 institutes, 885 studies were identified. An additional search of open-accessible databases, Academica Sinica, PubMed (NLM), Web of Science (AHCI), Web of Science (Core Collection), Web of Science (SCI) and Web of Science (SSCI), was performed as well. The search results are summarized in Table 4.

Table 4. Number of studies identified at search of currently available databases

	Database					
Keyword	AS	NLM	AHCI	CC	SCI	SSCI
autism[a] + exercise[b]	2	135	1	112	62	88
autism[a] + intervention	125	2312	23	3671	1158	3338
autism[a] + treatment	156	9744	17	4558	2656	3073
autism[a] + exercise[b] + intervention	1	23	0	30	9	28
autism[a] + exercise[b] + treatment	1	72	0	19	12	13

[a]autism: autism; autism spectrum disorder; ASD
[b]exercise: exercise; fitness; physical activity; aerobic; motor function; motor skill; motor performance

From Academica Sinica, 285 studies were identified. Among the 285 studies, 10 remained after screening of the abstracts. Among the 10 studies, 2 were review articles and only 9 studies were relevant. From PubMed (NLM), 12286 studies were identified. After screening the abstracts, 21 studies were selected/remained. From Web of Science (AHCI), 41 studies were identified. After screening the abstracts, only 1 study was selected/remained. From Web of Science (Core Collection), 8390 studies were identified. After screening the abstracts, 15 studies were selected/remained. From Web of Science

(SCI), 3896 studies were identified. After screening the abstracts, 9 studies were selected/remained. From Web of Science (SSCI), 6540 studies were identified. After screening the abstracts, 13 studies were selected/remained. Details of the inclusion procedure please refer to the Figure 3 below.

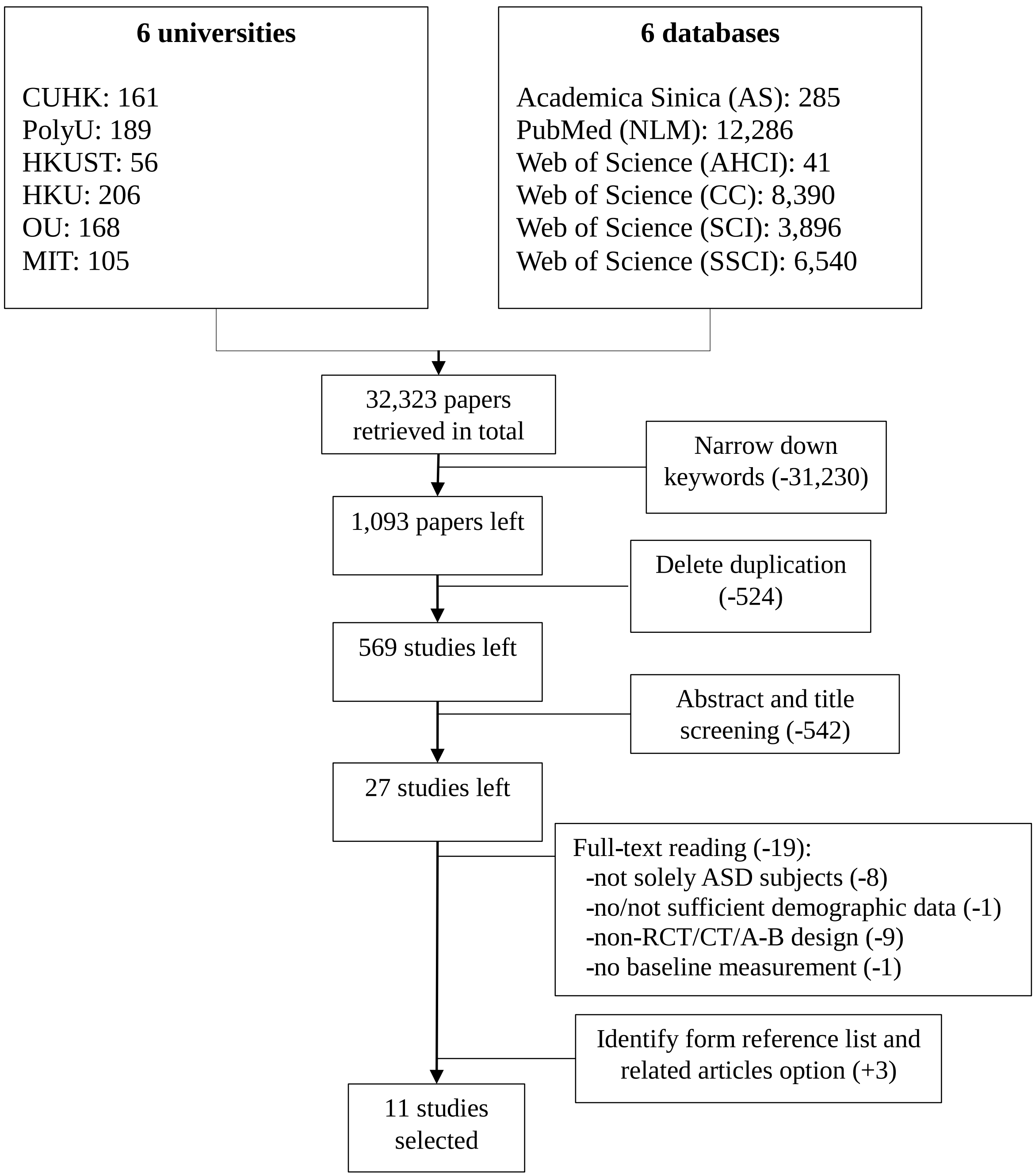


Figure 3. Flowchart of the study inclusion

Comply the full list of included studies

From the initial search, 32,323 papers were retrieved in total. While the keywords were narrowed down as “autism” OR “autism spectrum disorder” OR “ASD” AND (“exercise” OR “fitness” OR “physical activity” OR “aerobic” OR “motor function” OR “motor skill” OR “motor performance”) AND (“intervention” OR “treatment”), 1,093 papers were left. After deleting duplications, 569 studies were identified. 27 studies then left after title and abstract screening (441 dropped by title screen; 101 dropped by abstract screen in details).

After reading all the full-texts of those 27 studies, 19 studies were excluded because of insufficient information presented (details please refer to figure 3 above). An additional 3 were selected in the later part of the procedure: one study (Wuang, Wang, Huang, & Su, 2010) was identified from the reference list of the remained articles; two studies (Kern, Fletcher, Garver, Mehta, Grannemann, Knox, & Trivedi, 2011; Koenig, Buckley-Reen, & Garg, 2012) were identified using the “related articles” option from the Web of Science. The total number of studies identified was 11 (27-19+1+2=11).

The selected 11 studies were reviewed by a second reviewer (an educational psychologist) for confirming the appropriateness of inclusion. The details of characteristics of the included studies were listed in Table 4 and 5. The measures and findings of the included studies were shown in Table 6. Quality assessment were conducted for these 11 identified studies, as for inclusion in this systematic review. The results of the quality assessment were revealed in Table 8 (Jadad) and 9 (CONSORT).

Characteristics of the included studies

Participants

A total 303 participants were involved in the eleven included studies (Bahrami, Movahedi, Marandi, & Abedi, 2012; Bass, Duchowny, & Llabre, 2009; Chan, Sze, Siu, Lau, & Cheung, 2013; Fragala-Pinkham, Haley, & O'Neil, 2008; Fragala-Pinkham, Haley, & O’Neil, 2011; Kern et al., 2011; Koenig, Buckley-Reen, & Garg, 2012; Pan, 2011; Pan, 2010; Pitetti, Rendoff, Grover, & Beets, 2007; Wuang et al., 2010). The sample sizes of eleven included studies ranged from 10 to 60. Sexual distribution was reported in all the 11 studies. The ratio of male participants in the included studies was 83.17%. The proportions of males in each study ranged from 60% to 100%. The average age of the participants reported by the eleven included studies was from 7.2 to 17.4 years old (see Table 5).

Table 5. Characteristics of included studies

Study	**N** (Male, Female)	**Age** (M±SD)	**Place**
Bahrami et al. (2012)	30 (26, 4)	5-16 (T: 9.20±3.32; C: 9.06±3.33)	Iran
Bass et al. (2009)	34 (29, 5)	4-10 (T: 6.95±1.67; C: 7.73±1.65)	US
Chan et al. (2013)	40 (36, 4)	6-17 (T: 11.28±3.90; C: 12.42±3.25)	HK
Fragala-Pinkham et al. (2008)	16 (11, 5)	6-12 (9.58±1.33)	US
Fragala-Pinkham et al. (2011)	12 (11, 1)	6-12 (T: 9.60±2.60; C: 9.60±1.30)	US
Kern et al. (2011)	24 (18, 6)	3-12 (7.80±2.9)	US
Koening et al. (2012)	46 (37, 9)	5-12 (T: 9.58; C: 8.58)	US
Pan (2010)	16 (16, 0)	6-9 (T: 7.27±1.25; C: 7.20±.89)	Taiwan
Pan (2011)	15 (15, 0)	7-12 (T: 9.31±1.67; C: 8.75±1.76)	Taiwan
Pitetti et al. (2007)	10 (6, 4)	14-19 (T: 16.60±1.90; C: 17.40±1.10)	US
Wuang et al. (2010)	60 (47, 13)	6-10 (M: 9.10±4.11; F: 9.11±3.76)	Taiwan

Table 5. Characteristics of included studies (continue)

Study	**N** (Control)	**Design**	**Intervention**
Bahrami et al. (2012)	30 (15)	RCT	Kata techniques training (56sessions, 90 min/session; 4day/week,)
Bass et al. (2009)	34 (15)	RCT	Horse riding (1hr/session, 1session/week; 12 weeks)
Chan et al. (2013)	40 (19)	RCT	Yoga (1hr/session, 2session/week; 4weeks)
Fragala-Pinkham et al. (2008)	16 (9)	A-B	Aquatic exercise (32-50min/session;2session/week; 14weeks)
Fragala-Pinkham et al. (2011)	12 (5)	CT	Aquatic exercise (40min/session; 2sessions/week; 14weeks)
Kern et al. (2011)	24 (24)	A-B	Horse riding (1hr/session; 1session/week; 6months)
Koening et al. (2012)	46 (18)	CT	Yoga (15-20 min/session; 5sessios/week; 16weeks)
Pan (2010)	16 (8)	CT	Aquatic exercise (20sessions; 90min/session; 2sessions/week)
Pan (2011)	15 (8)	CT	Aquatic exercise (28sessions; 60min/session; 2sessions/week)
Pitetti et al. (2007)	10 (5)	CT	Treadmill walking (individualized progression program; 9months)
Wuang et al. (2010)	60 (30)	CT	Horse riding (1hr/session; 2sessions/week; 40sessions)

Venues

Six studies (Bass et al., 2009; Fragala-Pinkham, et al., 2008; Fragala-Pinkham et al., 2011; Kern et al., 2011; Koenig et al., 2012; Pitetti et al., 2007) were conducted in the United States, three studies were conducted in Taiwan; and the remaining two were conducted in Iran and Hong Kong.

Among all 303 participants, 147 were in experimental group and the rest of them (156) were in control group. Regarding to the study design, 3 studies (Bahrami et al., 2012;

Bass et al., 2009; Chan et al., 2013) adopted the design of randomized control trial; 6 studies (Fragala-Pinkham et al., 2011; Koenig et al., 2012; Pan, 2011; Pan, 2010; Pitetti et al., 2007; Wuang et al., 2010) adopted control trial; and 2 studies (Fragala-Pinkham et al., 2008; Kern et al., 2011) used the A-B crossover design.

Types of exercise programs

Among the eleven included studies, five types of exercises were adopted as the intervention programs, those were Aquatic exercise (4 studies), Horse riding (3 studies), Kata (1 study), Treadmill walking/jogging (1 study), and Yoga (2 studies).

Duration and frequency

The frequency of those exercise programs varied from once a week (Bass et al., 2009; Kern et al., 2011) to five times a week (Koenig et al., 2012), duration varied from 15 minutes to 90 minutes. Details were shown in Table 5 below.

Measures

Sixty seven measures were used by the eleven included studies for evaluating the outcome of exercise interventions. Table 5 summarized the characteristics of the measures and the details were described as below.

In Bahrami et al. (2012)'s study, the measure Stereotypy Subscale of Gilliam Autism Rating Scale - 2nd Edition (SS-GARS-2) was adopted, it has 14 items in total; administrated to parents/teachers; and scaled from 0 (never) to 3 (always).

In Bass et al. (2009)'s study, two scales - Sensory Profile (SP) and Social Responsiveness Scale (SRS), were adopted; SP has 125 items, administrated to parents/teachers, and scaled from 1 (always) to 5 (never); SRS has 65 items, administrated to parents/teachers, and scaled from 0 (never) to 3 (always).

In Chan et al. (2013)'s study, four checklists were adopted, which included 1) Autism Treatment Evaluation Checklist - Sensory or Cognitive (SEN-ATEC) [18-item; by parents/teachers; 1 (not true) to 3 (very true)]; 2) Autism Treatment Evaluation Checklist - Health or Physical (PHY-ATEC) [25-item; by parents/teachers; 1 (not a problem) to 4 (serious problem)]; 3) Autism Treatment Evaluation Checklist - Sociability (SOC-ATEC) [20-item; by parents/teachers; 1 (not descriptive) to 3 (very descriptive)]; and 4) Autism Treatment Evaluation Checklist - Speech or Language or Communication (SPE-ATEC) [14-item; by parents/teachers; 1 (not true) to 3 (very true)].

In Fragala-Pinkham et al. (2008)'s study, one inventory, three physical tests and four flexor/extensor measures were adopted, they were Multidimensional Pediatric Evaluation of Disability Inventory (M-PEDI) [159-item; by parents; scale score ranging from 0 to 100], Floor to stand test - FTS (sec) [participant is timed while getting up off the floor, walking 3 meters, and returning to a sitting position on the floor; by pediatric physical therapists], Half mile run (min) [participant is timed while doing a half mile run; by pediatric physical therapists], Modified curl-up (count) [abdominal muscular strength and endurance measurement as specified in the Brockport Fitness Test Manual; by pediatric physical therapists], Hip abductors (kg), Knee extensors (kg), Knee flexors (kg) and Ankle plantarflexors (kg) [peak isometric muscle strength measured with a handheld Chatillon dynamometer using previously established protocols; by pediatric physical therapists].

In Fragala-Pinkham et al. (2011)'s study, one scale, one mobility test and three physical fitness tests were adopted. Those included Swimming Classification Scale (levels) [16-item; by parents; levels ranging 1 to 5], Mobility skills (scaled scores) test [159-item; by parents; scale score ranging from 0 to 100]; Half mile run (min) [Participant is timed while doing a half mile run; by pediatric physical therapists]; Isometric Push-Up - IPU (sec) [a muscle endurance test by isometric push-up (seconds) according to the Brockport Fitness Test manual]; and Modified curl-ups (repetitions) [muscle endurance measurement as specified in the Brockport Fitness Test Manual].

In Kern et al. (2011)'s study, three scales were adopted - Childhood Autism Rating Scale

(CARS), Timberlawn Parent-Child Interaction Scale (TPCIS), and Sensory Profile (SP); CARS has 15 items, administrated by research assistant, and its scale score ranged from 15 to 60; TPCIS has 6 sub-scales, and was a 5-point Likert scale; SP has 125 items, administrated to parents/teachers; ranged from 1 (always) to 5 (never).

The measure Aberrant Behavior Checklist (ABC) - Community was adopted by Koening (2012), it has 58 items; administrated to parents/teachers; and was a 4-point scale ranging from 0 (not a problem) to 3 (a severe problem).

Humphries Assessment of Aquatic Readiness (HAAR) and the Social competence and Antisocial behavior under the School Social Behavior Scale (SSBS-2) were used in Pan (2010) study, HAAR was a multidimensional water skill test (i.e., balance and controlled movement, rotations, and etc.), whereas SSBS-2 was a 32-item scale; administrated by teachers, its scale score ranged from 32 to 160.

HAAR was also adopted by Pan in her 2011 study. On the other hand, three physical tests - Sit and Reach (SnR), Curl-ups (30s and 60s), and the Progressive Aerobic Cardiovascular Endurance Run (PACER) [16-m distance; participants run as long as possible back and forth across a 16-m distance at a specified pace, which gets faster each minute, until the children could no longer maintain], two body fat measurements - Bioelectrical impedance analysis (BIA) (MF-BIA8, InBody 720, Biospace) [participants stood on the footplate with bare feet and held both hand electrodes for 2 min] and Body mass index (BMI) [body mass (kg) divided by the square of their height (m)], were adopted in Pan (2011)'s study as well.

The measure BMI was also adopted by Pitetti et al. (2007), same as Pan (2011) study, it was calculated as dividing participant's body mass (kg) by the square of their height (m).

The Bruininks-Oseretsky Test of Motor Proficiency (BOTMP) was adopted in Wuang et al. (2010) study, it was designed for children ranging in age from 4.5 to 14.5 years of age; the test consists of separate measures of gross, fine motor skills and upper-limb coordination; the higher the BOTMP composite scores, the better the motor outcome. The Test of Sensory Integration Function (TSIF) was also adopted by Wuang et al. (2010), it was designed for children aged from 3 to 12 years; consists of 98 items; divided into 7 subtests: postural-ocular movement, bilateral integration, sensory discrimination, sensory modulation, sensory searching, attention and activity, and emotion and behavior. TSIF was originally a 5-point Likert scale (see Table 6).

Table 6. Characteristics of outcome measures

Study	Measure	
Bahrami et al. (2012)	Stereotypy Subscale of Gilliam Autism Rating Scale - 2nd Edition (SS-GARS-2)	14-item; by parents/teachers; 0 (never) to 3 (always)
Bass et al. (2009)	Sensory Profile (SP)	125-item; by parents/teachers; 1 (always) to 5 (never)
	Social Responsiveness Scale (SRS)	65-item; by parents/teachers; 0 (never) to 3 (always)
Chan et al. (2013)	Autism Treatment Evaluation Checklist - Sensory or Cognitive (SEN-ATEC)	18-item; by parents/teachers; 1 (not true) to 3 (very true)
	Autism Treatment Evaluation Checklist - Health or Physical (PHY-ATEC)	25-item; by parents/teachers; 1 (not a problem) to 4 (serious problem)
	Autism Treatment Evaluation Checklist - Sociability (SOC-ATEC)	20-item; by parents/teachers; 1 (not descriptive) to 3 (very descriptive)
	Autism Treatment Evaluation Checklist – Speech or Language or Communication (SPE-ATEC)	14-item; by parents/teachers; 1 (not true) to 3 (very true)
Fragala-Pinkham et al. (2008)	Multidimensional Pediatric Evaluation of Disability Inventory (M-PEDI)	159-item; by parents; scale score ranging from 0 to 100

Table 6. Characteristics of outcome measures (continue)

Study	Measure	
Fragala-Pinkham et al. (2008)	Floor to stand - FTS (sec)	Participant is timed while getting up off the floor, walking 3 meters, and returning to a sitting position on the floor; by pediatric physical therapists
	Half mile run (min)	Participant is timed while doing a half mile run; by pediatric physical therapists
	Hip abductors (kg)	Peak isometric muscle strength measured with a handheld Chatillon dynamometer using previously established protocols; by pediatric physical therapists
	Knee extensors (kg)	Peak isometric muscle strength measured with a handheld Chatillon dynamometer using previously established protocols; by pediatric physical therapists
	Knee flexors (kg)	Peak isometric muscle strength measured with a handheld Chatillon dynamometer using previously established protocols; by pediatric physical therapists
	Ankle plantarflexors (kg)	Peak isometric muscle strength measured with a handheld Chatillon dynamometer using previously established protocols; by pediatric physical therapists
	Modified curl-up (count)	Abdominal muscular strength and endurance measurement as specified in the Brockport Fitness Test Manual; by pediatric physical therapists
Fragala-Pinkham et al. (2011)	Swimming Classification Scale (levels)	16-item; by parents; levels ranging 1 to 5
	Mobility skills (scaled scores)	159-item; by parents; scale score ranging from 0 to 100

Table 6. Characteristics of outcome measures (continue)

Study	Measure	
Fragala-Pinkham et al. (2011)	Half mile run (min)	Participant is timed while doing a half mile run; by pediatric physical therapists
	Isometric Push-Up - IPU (sec)	A muscle endurance test by isometric push-up (seconds) according to the Brockport Fitness Test manual
	Modified curl-ups (repetitions)	Muscle endurance measurement as specified in the Brockport Fitness Test Manual
Kern et al. (2011)	Childhood Autism Rating Scale (CARS)	15-item; by research assistant; scale score ranging from 15 to 60
	Timberlawn Parent-Child Interaction Scale (TPCIS)	By research assistant; 5 point scale on each domain
	TPCIS - Expressiveness TPCIS - Responsiveness TPCIS - Positive Regard TPCIS - Negative Regard TPCIS - Mood and Tone TPCIS - Empathy	
	Sensory Profile (SP)	125-item; by parents/teachers; 1 (always) to 5 (never)
	SP-Auditory high-threshold SP-Auditory low-threshold SP-Visual high-threshold SP-Visual low-threshold SP-Touch high-threshold SP-Touch low-threshold SP-Vestibular high-threshold SP-Vestibular low-threshold	

Table 6. Characteristics of outcome measures (continue)

Study	Measure	
Koening et al. (2012)	Aberrant Behavior Checklist (ABC) - Community	58-item; by parents/teachers; 4-point scale ranging from 0 (not a problem) to 3 (a severe problem)
Pan (2010)	Humphries Assessment of Aquatic Readiness -IV - Balance and controlled movement (HAAR-IV)	8-item; 0 (unable) to 1 (able)
	Humphries Assessment of Aquatic Readiness -II – Introduction to water environment (HAAR-II)	10- items; 0 (unable) to 1 (able)
	Humphries Assessment of Aquatic Readiness -III – Rotations (HAAR-III)	3- items; 0 (unable) to 1 (able)
	Humphries Assessment of Aquatic Readiness -V – Independent movement in water (HAAR-V)	6- items; 0 (unable) to 1 (able)
	Social competence -School Social Behavior Scales (SC-SSBS-2)	32-item; by teachers, scale score ranging from 32 to 160
	Antisocial behavior -School Social Behavior Scales (AnB-SSBS-2)	32-item; by teachers, scale score ranging from 32 to 160

Table 6. Characteristics of outcome measures (continue)

Study	Measure	
Pan (2011)	Humphries Assessment of Aquatic Readiness -IV - Balance and controlled movement (HAAR-IV)	8-item; 0 (unable) to 1 (able)
	Humphries Assessment of Aquatic Readiness -III – Rotations (HAAR-III)	3- items; 0 (unable) to 1 (able)
	Humphries Assessment of Aquatic Readiness -V – Independent movement in water (HAAR-V)	6- items; 0 (unable) to 1 (able)
	Sit and Reach (SnR) Physical fitness-SnR	A flexibility test of the hamstring muscles and lower back from the Taiwan Ministry of Education Physical Fitness Test manual
	Physical fitness-Percent body fat	Bioelectrical impedance analysis (BIA) (MF-BIA8, InBody 720, Biospace); participants stood on the footplate with bare feet and held both hand electrodes for 2 min.
	Physical fitness-Curl-ups (30s)	Curl-up from the Taiwan Ministry of Education Physical Fitness Test; participants lie in a supine position on a mat with their knees bent at about 140 degree and feet flat on the floor. Their hands were placed on the front of the thighs rather than on the mat alongside the body. As the participant curls up, the hands slide along the thighs until the fingertips contact the patellae.

Table 6. Characteristics of outcome measures (continue)

Study	**Measure**	
Pan (2011)	Physical fitness-Curl-ups (60s)	Curl-up from the Taiwan Ministry of Education Physical Fitness Test.
	The Progressive Aerobic Cardiovascular Endurance Run (PACER) Physical fitness-16-m PACER	16-m distance; participants run as long as possible back and forth across a 16-m distance at a specified pace, which gets faster each minute, until the children could no longer maintain.
	Body mass index (BMI) Physical fitness-BMI	Participant’s body mass (kg) divided by the square of their height (m)
Pitetti et al. (2007)	Body mass index (BMI)	Participant’s body mass (kg) divided by the square of their height (m)
Wuang et al. (2010)	Bruininks-Oseretsky Test of Motor Proficiency -Gross Motor (BOTMP-GM) Running speed and agility Balance Bilateral coordination Strength	For children ranging in age from 4.5 to 14.5 years of age; the test consists of separate measures of gross, fine motor skills and upper-limb coordination; the higher the BOTMP composite scores, the better the motor outcome.

Table 6. Characteristics of outcome measures (continue)

Study	Measure	
Wuang et al. (2010)	Bruininks-Oseretsky Test of Motor Proficiency - Fine Motor (BOTMP-FM) Response speed Visual-motor control Upper-limb speed and dexterity	For children ranging in age from 4.5 to 14.5 years of age; the test consists of separate measures of gross, fine motor skills and upper-limb coordination; the higher the BOTMP composite scores, the better the motor outcome.
	Bruininks-Oseretsky Test of Motor Proficiency (BOTMP) - Upper-limb coordination	For children ranging in age from 4.5 to 14.5 years of age; the test consists of separate measures of gross, fine motor skills and upper-limb coordination; the higher the BOTMP composite scores, the better the motor outcome.
	Test of Sensory Integration Function (TSIF) Postural-ocular movement Bilateral integration Sensory discrimination Sensory modulation Sensory searching Attention and activity Emotion and behavior	For children aged from 3 to 12 years; consists of 98 items; divided into 7 subtests: postural-ocular movement, bilateral integration, sensory discrimination, sensory modulation, sensory searching, attention and activity, and emotion and behavior; a 5-point Likert scale from 1 (*never*) to 5 (*always*)

Findings

Sixty-seven measures were used by the eleven included studies for evaluating the outcome of exercise interventions. Among the 62 measures, 53.73% (36 out of 67) were detected a slightly higher, with the p-value less than .05, in treatment groups compared to the control groups. Among them, 75% (27 out of 36) found significant results, with p-value less than .01, in the difference of outcome variables between treatment and control groups (see Table 7).

Table 7. Measures and findings of included studies

Study	Measure	Finding
Bahrami et al. (2012)	SS-GARS-2	T > C**
Bass et al. (2009)	SP	T > C**
	SRS	T > C*
Chan et al. (2013)	SEN-ATEC	T > C*
	PHY-ATEC	T > C**
	SOC-ATEC	T > C**
	SPE-ATEC	N.S.
Fragala-Pinkham et al. (2008)	M-PEDI	N.S.
	FTS (sec)	N.S.
	Half mile (min)	T > C**
	Hip abductors (kg)	N.S.
	Knee extensors (kg)	N.S.
	Knee flexors (kg)	N.S.
	Ankle plantarflexors (kg)	N.S.
	Modified curl-up (count)	N.S.
Fragala-Pinkham et al. (2011)	Swimming Classification Scale	T > C*
	(levels)	N.S.
	Mobility skills (scaled scores)	N.S.
	1/2 Mile (min)	N.S.
	IPU (sec)	N.S.
	Modified curl-ups (repetitions)	
Kern et al. (2011)	TPCIS-Expressiveness	N.S.
	TPCIS -Responsiveness	N.S.
	TPCIS -Positive Regard	N.S.
	TPCIS -Negative Regard	N.S.
	TPCIS -Mood and Tone	T > C**
	TPCIS -Empathy	N.S.

*p < .05; **p < .01

Table 7. Measures and findings of included studies (continue)

Study	Measure	Finding
Kern et al. (2011)	SP-Auditory high-threshold	T > C*
	SP-Auditory low-threshold	N.S.
	SP-Visual high-threshold	N.S.
	SP-Visual low-threshold	N.S.
	SP-Touch high-threshold	N.S.
	SP-Touch low-threshold	N.S.
	SP-Vestibular high-threshold	N.S.
	SP-Vestibular low-threshold	N.S
	CARS	T > C*
Koening et al. (2012)	ABC-Community-maladaptive behavior	T > C*
Pan (2010)	HAAR-II	T > C**
	HAAR-III	T > C**
	HAAR-IV	T > C**
	HAAR-V	T > C**
	SC-SSBS-2	N.S.
	AnB-SSBS-2	T > C**
Pan (2011)	HAAR-IV	T > C**
	HAAR-III	N.S.
	HAAR-V	T > C*
	Physical fitness-SnR	N.S.
	Physical fitness-Percent body fat	N.S.
	Physical fitness-Curl-ups (30s)	T > C*
	Physical fitness-Curl-ups (60s)	T > C*
	Physical fitness-16-m PACER	N.S.
	Physical fitness-BMI	N.S.
Pitetti et al. (2007)	BMI	N.S.
Wuang et al. (2010)	BOTMP-Running speed and agility	T > C**
	BOTMP-Balance	T > C**
	BOTMP-Bilateral coordination	T > C**
	BOTMP-Strength	T > C**
	BOTMP-Upper-limb coordination	T > C**
	BOTMP-Response speed	T > C**
	BOTMP-Visual-motor control	T > C**
	BOTMP-Upper-limb speed and dexterity	T > C**
	TSIF-Postural-ocular movement	T > C**
	TSIF-Bilateral integration	T > C**
	TSIF-Sensory discrimination	T > C**
	TSIF-Sensory modulation	T > C**
	TSIF-Sensory searching	T > C**
	TSIF-Attention and activity	T > C**
	TSIF-Emotion and behavior	T > C**

*p < .05; **p < .01

Results of the quality assessment

The results of the critical appraisal were listed in table 8 (by the appraisal tool, Jadad) and table 9 (by the appraisal tool, modified CONSORT).

All the data extraction as well as the quality assessment procedures were performed by two individual raters; discrepancies were solved by consensus.

In table 8, most of the studies (54.54%; 6 out of 11) got only 1 mark according to the Jadad critical appraisal criteria; 4 of them (36.36%) got 3 marks; one study (Chan et al., 2013) got full marks.

Table 8. Critical appraisal of study quality - Jadad

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Bahrami et al. (2012)	1	1	0	0	1	0	0
Bass et al. (2009)	1	1	0	0	1	0	0
Chan et al. (2013)	1	1	1	1	1	0	0
Fragala-Pinkham et al. (2008)	0	0	0	0	1	0	0
Fragala-Pinkham et al. (2011)	0	0	0	0	1	0	0
Kern et al. (2011)	0	0	0	0	1	0	0
Koening et al. (2012)	0	0	0	0	1	0	0
Pan (2010)	0	0	1	1	1	0	0
Pan (2011)	0	0	1	1	1	0	0
Pitetti et al. (2007)	0	0	0	0	1	0	0
Wuang et al. (2010)	0	0	0	0	1	0	0

Q1: Randomization sounded (0/1)
Q2: Randomization procedure described (0/1)
Q3: Double-blinded (0/1)
Q4: Double-blinded procedure described (0/1)
Q5: Withdrawals/dropouts described (0/1)
Q6: Inappropriate randomization process (0/-1)
Q7: Inappropriate double-blinded process (0/-1)

In table 9, all of the 11 selected studies were over 7 (maximum 10) marks; 3 of them got 10 full marks; 6 of them got 9 marks; and the rest of 2 got 8 marks.

Table 9. Critical appraisal of study quality - CONSORT

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Bahrami et al. (2012)	1	1	1	1	1	1	1	1	1	1
Bass et al. (2009)	1	1	1	1	1	1	1	1	1	1
Chan et al. (2013)	1	1	1	1	1	1	1	1	1	1
Fragala-Pinkham et al. (2008)	0	1	1	1	1	0	1	1	1	1
Fragala-Pinkham et al. (2011)	1	1	1	1	1	0	1	1	1	1
Kern et al. (2011)	0	1	1	1	1	0	1	1	1	1
Koening et al. (2012)	1	1	1	1	1	0	1	1	1	1
Pan (2010)	1	1	1	1	1	0	1	1	1	1
Pan (2011)	1	1	1	1	1	0	1	1	1	1
Pitetti et al. (2007)	1	1	1	1	1	0	1	1	1	1
Wuang et al. (2010)	1	1	1	1	1	0	1	1	1	1

Q1: Trial design (0/1)
Q2: Participants (0/1)
Q3: Sample size (0/1)
Q4: Interventions (0/1)
Q5: Outcomes (0/1)
Q6: Randomization (0/1)
Q7: Statistical methods (0/1)
Q8: Primary objective (0/1)
Q9: Program dosage (0/1)
Q10: Program variables (0/1)

CHAPTER 3

Meta-analysis

Meta-analysis is a statistical technique for summarizing, and reviewing previously published quantitative research (Hedges & Olkin, 2014). It combines intervention/treatment effects into a common metric that is standardized, called an effect size. It can be used to compare different studies. In the case of intervention research, a meta-analysis may examine the outcomes from multiple interventions across studies and determine their relative effectiveness. The computation of effect size for various outcomes within the studies makes it possible to compare the characteristics of both the interventions and the participants.

Methods of the meta-analysis

Effect size

In this meta-analysis, the effect size, Cohen's d was adopted. This statistic represents the mean difference between the interventional group and control group measured in standard deviation unit. The effect size statistic provides information about both the direction and magnitude of treatment effect. Cohen's d values were computed, using Comprehensive Meta-Analysis software, version 2.2.064 (2011), by subtracting the control group mean from the interventional group mean and dividing this value by pooled sample standard deviation. However, according to Cohen 1988, the studies with small sample sizes might lead to overestimate of effect size values. This weakness was removed in our study by multiplying Cohen's d value by a coefficient that includes information on the sample sizes of experimental and control groups, resulting in the unbiased Cohen's effect size d. The effect size values of .20, .50 and .80 correspond with small, medium and large effects (see Table 10). The calculation of Cohen's effect size d was shown as below (Cohen, 1988).

$$d = \frac{\overline{x_t} - \overline{x_c}}{\sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}}$$

$\overline{x_t}$: mean of treatment group
$\overline{x_c}$: mean of control group
s_t^2: variance of treatment group
s_c^2: variance of control group
n_t: number of sample in treatment group
n_c: number of sample in control group

Table 10. Cohen's *d* effect size interpretation

Effect size	Standard
.8 or greater	Large
.7	
.6	
.5	Medium
.4	
.3	
.2	Small
.1	
.0	

Note. Adapted from Cohen (1988).

Model selection with Cochran Q

In computing the effect sizes, the post intervention data from the first available follow-up in the case of all outcomes were chose. The choice for this approach, because the exercise effects for these outcomes tend to be strongest soon after the intervention and gradually dissipate over time. After calculating the Cohen's effect sizes of individual studies for each outcome, these effect sizes were combined to give a mean Cohen's effect size d by means of fixed effect model or random effect model. The d value was given a positive sign when the intervention group did better on the outcome than the control group and a negative sign was assigned when the reverse was true. To examine whether the studies contributing to the mean Cohen's effect sizes d were come from a common population, the heterogeneity indicator, Cochran Q was computed. Cochran Q follows a chi-squared distribution with k-1 degrees of freedom, where k represents the number of independent effect sizes included. A statistically significant Q value indicates that there is an extensive amount of variability across the included studies. The calculation of Q was shown as below (Cochran, 1954).

$$Q = \sum w_i (T_i - \bar{T})^2$$

where w_i is the weighting factor for the ith study assuming a fixed-effects model, T_i is the usual estimate of a mean effect size for ith study, and $\bar{T}$ is defined as:

$$\bar{T} = \frac{\sum w_i T_i}{\sum w_i}$$

When the studies show no heterogeneity between them, it is appropriate to use fixed effect model in combining the results. On the other hand, when heterogeneity occurs among the studies, a random effect model should be used instead. The limitation of random effect model is that the statistical tests lack power and fail to reject the null hypothesis of homogeneous results even if substantial differences between studies exist.

Publication bias

A potential threat to the validity of a meta-analysis is publication bias. Publication bias has the potential to skew the results of the analysis due to potential sampling bias. Because published literature is more easily accessible to researchers. Most meta-analyses rely heavily on published literature. The problem inherent in sampling from published literature is that research the gets published is more likely to have found a positive results than research that report non-significant results. This bias has been described as a file-drawer bias because many research reports with non-significant findings are not published and sit instead in the researcher’s file drawer. This threat has the potential to have an upward bias, as more positive studies are included in the analysis and fewer non-significant or small effect studies are excluded, the analysis may report an inflated overall effect. An additional threat to the validity of a meta-analysis is the likelihood that no matter how exhaustive the literature search, that some studies will not be found and therefore excluded from the analysis.

Assessment of publication bias is conducted by plotting effect size by the standard error (or 1/standard error) of the studies on a funnel plot. This visual representation of the studies should look like an inverted funnel, with studies clustered around the mean symmetrically. If the studies are evenly distributed in an inverted funnel shape, it can be confirmed with confidence that there is not a systemic publication bias.

Procedure

Key procedures were undergone in this meta-analysis, which included the calculation of effect sizes, comparison of models, assessment of publication biases, and the moderation process (described above). Figure 4 summarized the procedures conducted in this study.

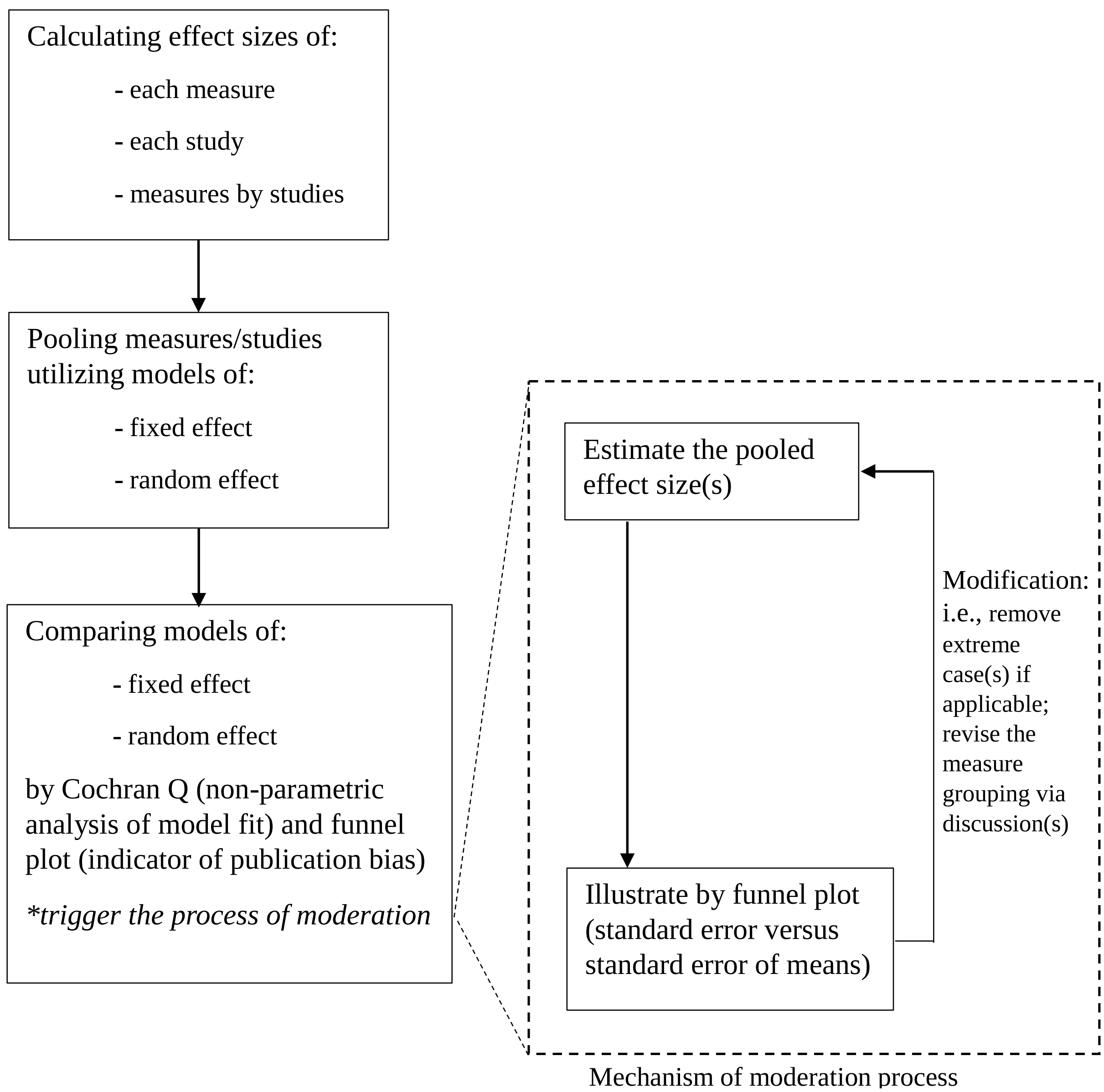


Figure 4. Overview of the meta-analysis procedure

Results of the meta-analysis

Eleven studies with 67 outcomes identified from the systematic review (Chapter 2) were included in this meta-analysis study.

Step 1: Effects by 67 measures and by 11 included studies

All 67 outcome measures extracted from those 11 selected studies, the overall effect size ($d = 1.535$, $p < .05$) showed a large effect with significance. While pooling those outcomes by studies, the overall effect size ($d = .998$, $p > .05$) showed still a large effect, but the result was not statistically significant.

Step 2: Categorization of 3 outcome domains

A further categorization with consensus of the expert team, across four professional fields (occupational therapy, educational psychology, speech therapy and physical education); three outcome categories were formed, which included Exercise Performance & Sport/Skill-related Fitness (EXE), Physiological & Biometric Indicator (PHY), and Social Cognition & Psychological Well-being (SOC). The outcome categorization procedure was similar to the previous study by Shin and Park (2012), which evaluated the effect of exercise programs for ID individuals, by lumping the outcome measures from over 15 to 4 outcome categories, including “Biometric and body composition”, “Exercise physiology”, “Measures of performance”, and “Professional/scholastic measures”.

Step 3: Effects by outcome domains/categories

The categorized outcome domains were undergone the analytical procedure for further investigation of the outcome effects. Throughout the exercise interventions, the outcome category of EXE showed a large effect but statistically insignificant ($d = 1.246$, $p > .05$), the PHY showed a large effect but also statistically insignificant ($d = 1.174$, $p > .05$), the only one showed a large effect with statistically significant was SOC ($d = 1.076$, $p < .05$).

Step 4: Identifying outliner(s)

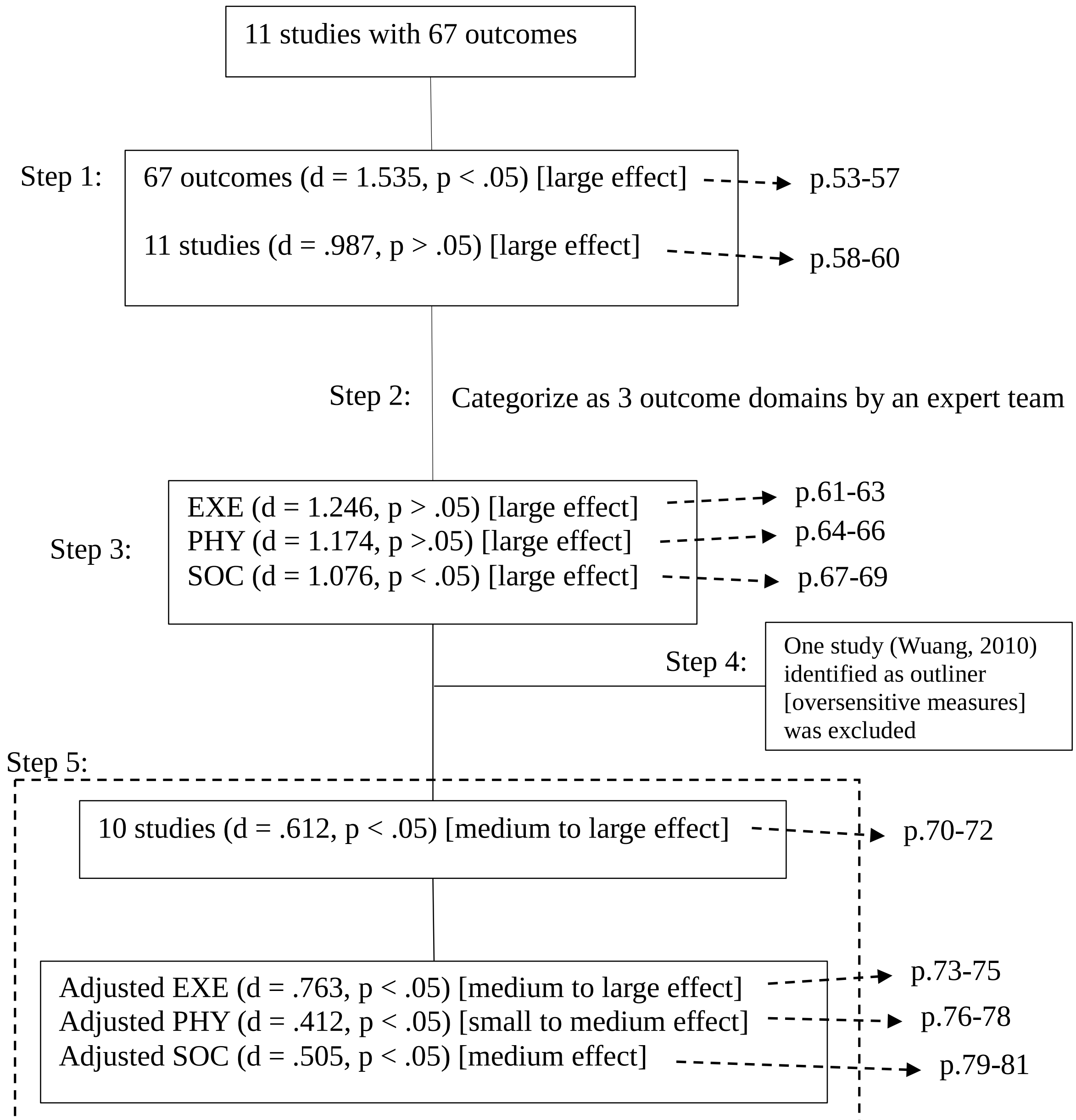
One study (Wuang et al., 2010)was identified as an outliner (those measures in this study were oversensitive to the intervention, which dominated the overall effect while combining outcome measures, regardless as by studies or by outcome domains), among all selected studies. In order to remove the effect from outliner(s), this study (Wuang, 2010) was finally decided to exclude for the later part of the analysis.

Step 5: Modifying the final list of included studies/measures on calculating the effect size

After excluding the outliner study (Wuang et al., 2010), the remaining 10 studies were underwent the analytical procedure again. By comparing the effects between 11 studies and 10 studies, an effect size calculation of the pooled outcomes by 10 studies was performed, and it showed a medium to large effect with significance ($d = .612$, $p < .05$), whereas the pooled outcomes by 11 studies showed a large effect without significance ($d = .998$, $p > .05$).

While excluding the outliner study (Wuang et al., 2010), the adjusted outcome categories were also undergone the analytical procedure for investigating the exercise effects for children with ASD. The EXE outcome category showed a medium to large effect with significance ($d = .763$, $p < .05$); the PHY category showed a small to medium effect with significance ($d = .412$, $p < .05$); and the SOC category showed a medium effect, also with statistical significance ($d = .505$, $p < .05$). Those adjusted outcome domains generally showed an average medium effect on the exercise interventions, and those effects were all statistically significant.

The summary of the step-by-step procedure and results was shown in Figure 5. Detailed results of each analysis were described below.



Note: Exercise performance & sport/skill-related fitness (EXE)
Physiological & biometric indicator (PHY)
Social cognition & psychological well-being (SOC)

Figure 5. Summary of procedure and results of the meta-analysis

All 67 measures within the 11 selected studies

In the figures 6 and 7, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.988, 95% CI = 0.891, 1.084, p = 0.000) showed the exercise gave a large effect on enhancing performance in all 67 measures among children with ASD. The funnel plots showed a right-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 8 and 9, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 1.535, 95% CI = 1.144, 1.926, p = 0.000) showed the exercise gave a large effect on enhancing performance in all 67 measures among children with ASD. The funnel plots showed a right-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

For the outcome category of all 67 measures, the heterogeneity Q statistic showed a statistical significance (Q = 1046.272, df (Q) = 66, p = 0.000), hence, a random effect model was more appropriate in combining the data.

Forest plot of outcome measures by standard difference (Std diff) in means and 95% CI

Study name	Outcome	Statistics for each study						
		Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Chan (2013)_04	SPE-ATEC	0.248	0.322	0.103	-0.383	0.878	0.770	0.442
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2008)_02	FTS	0.240	0.418	0.175	-0.579	1.060	0.575	0.566
Fragala-Pinkham (2008)_03	Half mile (min)	0.793	0.431	0.186	-0.053	1.638	1.837	0.066
Fragala-Pinkham (2008)_04	Hip abductors (kg)	0.177	0.417	0.174	-0.641	0.996	0.425	0.671
Fragala-Pinkham (2008)_05	Knee extensors (kg)	0.089	0.417	0.174	-0.728	0.906	0.214	0.830
Fragala-Pinkham (2008)_06	Knee flexors (kg)	0.226	0.418	0.175	-0.593	1.045	0.541	0.589
Fragala-Pinkham (2008)_07	Ankle plantarflexors (kg)	0.065	0.417	0.174	-0.752	0.882	0.156	0.876
Fragala-Pinkham (2008)_08	Modified curl-up (count)	0.412	0.421	0.177	-0.412	1.237	0.980	0.327
Fragala-Pinkham (2011)_01	Swimming Classification Scale (levels)	0.602	0.598	0.358	-0.571	1.774	1.006	0.315
Fragala-Pinkham (2011)_02	Mobility skills (scaled scores)	0.166	0.587	0.344	-0.983	1.316	0.284	0.777
Fragala-Pinkham (2011)_03	1/2 Mile	0.396	0.591	0.349	-0.762	1.555	0.671	0.502
Fragala-Pinkham (2011)_04	IPU	0.519	0.595	0.354	-0.647	1.686	0.873	0.383
Fragala-Pinkham (2011)_05	Modified curl-ups (repetitions)	0.252	0.588	0.345	-0.900	1.404	0.428	0.669
Kern (2011)_00	CARS	0.390	0.298	0.089	-0.194	0.974	1.308	0.191
Kern (2011)_01	TPCIS-Expressiveness	0.106	0.295	0.087	-0.472	0.685	0.360	0.718
Kern (2011)_02	TPCIS-Responsiveness	0.000	0.295	0.087	-0.579	0.579	0.000	1.000
Kern (2011)_03	TPCIS-Positive Regard	0.080	0.295	0.087	-0.499	0.658	0.270	0.787
Kern (2011)_04	TPCIS-Negative Regard	0.080	0.295	0.087	-0.499	0.659	0.271	0.786
Kern (2011)_05	TPCIS-Mood and Tone	0.197	0.296	0.088	-0.383	0.777	0.666	0.505
Kern (2011)_06	TPCIS-Empathy	0.063	0.295	0.087	-0.516	0.642	0.214	0.831
Kern (2011)_07	SP-Auditory high-threshold	0.467	0.299	0.089	-0.119	1.053	1.561	0.118
Kern (2011)_08	SP-Auditory low-threshold	0.011	0.295	0.087	-0.567	0.590	0.038	0.969
Kern (2011)_09	SP-Visual high-threshold	0.027	0.295	0.087	-0.552	0.605	0.091	0.928
Kern (2011)_10	SP-Visual low-threshold	0.110	0.295	0.087	-0.469	0.689	0.371	0.711
Kern (2011)_11	SP-Touch high-threshold	0.008	0.295	0.087	-0.571	0.586	0.026	0.980
Kern (2011)_12	SP-Touch low-threshold	0.070	0.295	0.087	-0.509	0.649	0.238	0.812
Kern (2011)_13	SP-Vestibular high-threshold	0.384	0.298	0.089	-0.200	0.968	1.289	0.197
Kern (2011)_14	SP-Vestibular low-threshold	0.283	0.297	0.088	-0.299	0.864	0.953	0.341
Koening (2012)_01	ABC-Community-maladaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_01	HAAR-IV	2.235	0.637	0.406	0.986	3.484	3.507	0.000
Pan (2010)_03	HAAR-II	2.445	0.661	0.437	1.150	3.741	3.700	0.000
Pan (2010)_04	HAAR-III	1.483	0.565	0.319	0.376	2.589	2.626	0.009
Pan (2010)_05	HAAR-V	3.276	0.765	0.585	1.776	4.775	4.282	0.000
Pan (2010)_06	SC-SSBS-2	0.633	0.512	0.263	-0.371	1.637	1.236	0.217
Pan (2010)_07	AnB-SSBS-2	0.926	0.526	0.277	-0.106	1.957	1.759	0.079
Pan (2011)_01	HAAR-IV	1.205	0.562	0.316	0.103	2.307	2.142	0.032
Pan (2011)_04	HAAR-III	0.978	0.547	0.300	-0.095	2.051	1.786	0.074
Pan (2011)_05	HAAR-V	1.293	0.569	0.324	0.178	2.408	2.273	0.023
Pan (2011)_06	Physical fitness-SnR	0.152	0.518	0.269	-0.863	1.168	0.294	0.769
Pan (2011)_07	Physical fitness-Percent body fat	0.513	0.526	0.277	-0.518	1.544	0.975	0.330
Pan (2011)_08	Physical fitness-Curl-ups (30s)	1.163	0.559	0.313	0.067	2.259	2.079	0.038
Pan (2011)_09	Physical fitness-Curl-ups (60s)	1.133	0.557	0.311	0.041	2.226	2.033	0.042
Pan (2011)_10	Physical fitness-16-m PACER	0.511	0.526	0.277	-0.520	1.542	0.972	0.331
Pan (2011)_11	Physical fitness-BMI	0.848	0.540	0.292	-0.211	1.907	1.570	0.116
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Wuang (2010)_01	BOTMP-Running speed and agility	4.422	0.479	0.230	3.483	5.361	9.228	0.000
Wuang (2010)_02	BOTMP-Balance	3.333	0.399	0.159	2.551	4.115	8.352	0.000
Wuang (2010)_03	BOTMP-Bilateral coordination	4.970	0.522	0.273	3.947	5.994	9.521	0.000
Wuang (2010)_04	BOTMP-Strength	5.129	0.535	0.286	4.081	6.177	9.593	0.000
Wuang (2010)_05	BOTMP-Upper-limb coordination	4.357	0.474	0.225	3.428	5.286	9.188	0.000
Wuang (2010)_06	BOTMP-Response speed	6.804	0.673	0.452	5.485	8.122	10.115	0.000
Wuang (2010)_07	BOTMP-Visual-motor control	8.086	0.782	0.611	6.553	9.618	10.340	0.000
Wuang (2010)_08	BOTMP-Upper-limb speed and dexterity	7.394	0.723	0.522	5.978	8.810	10.231	0.000
Wuang (2010)_09	TSIF-Postural-ocular movement	5.294	0.548	0.300	4.220	6.367	9.662	0.000
Wuang (2010)_10	TSIF-Bilateral integration	5.245	0.544	0.296	4.179	6.312	9.642	0.000
Wuang (2010)_11	TSIF-Sensory discrimination	5.355	0.553	0.306	4.271	6.438	9.686	0.000
Wuang (2010)_12	TSIF-Sensory modulation	5.270	0.546	0.298	4.200	6.340	9.652	0.000
Wuang (2010)_13	TSIF-Sensory searching	2.524	0.346	0.120	1.846	3.202	7.294	0.000
Wuang (2010)_14	TSIF-Attention and activity	4.406	0.478	0.228	3.469	5.343	9.219	0.000
Wuang (2010)_15	TSIF-Emotion and behavior	3.755	0.429	0.184	2.914	4.596	8.750	0.000
		0.988	0.049	0.002	0.891	1.084	19.980	0.000

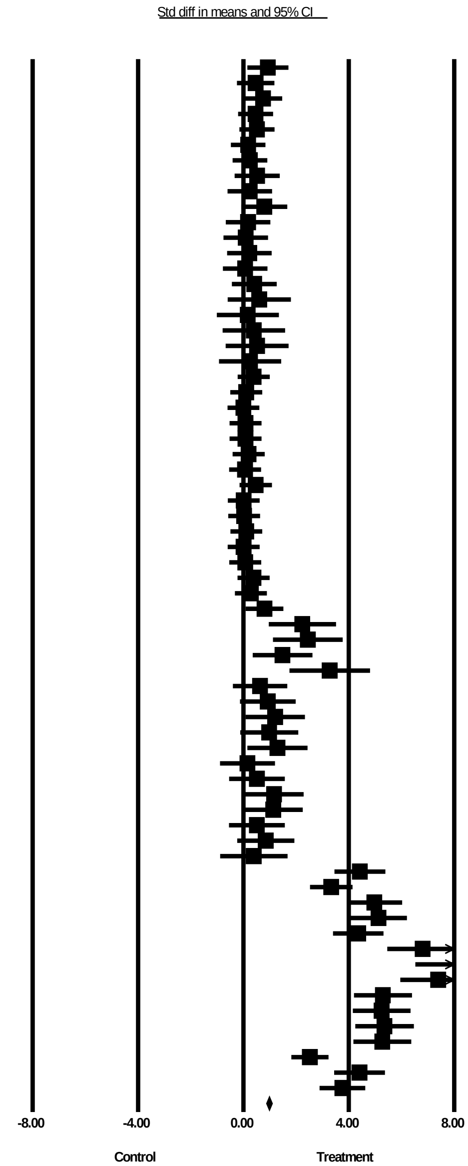


Figure 6. Forest Plot of all 67 outcomes – fixed effect model

*only the first author and publication year were shown; the underscore and number after the study name, i.e., the “_01” of “Bahrami (2012)_01”, indicates the first measure in the study; it applies to all tables and figures below.

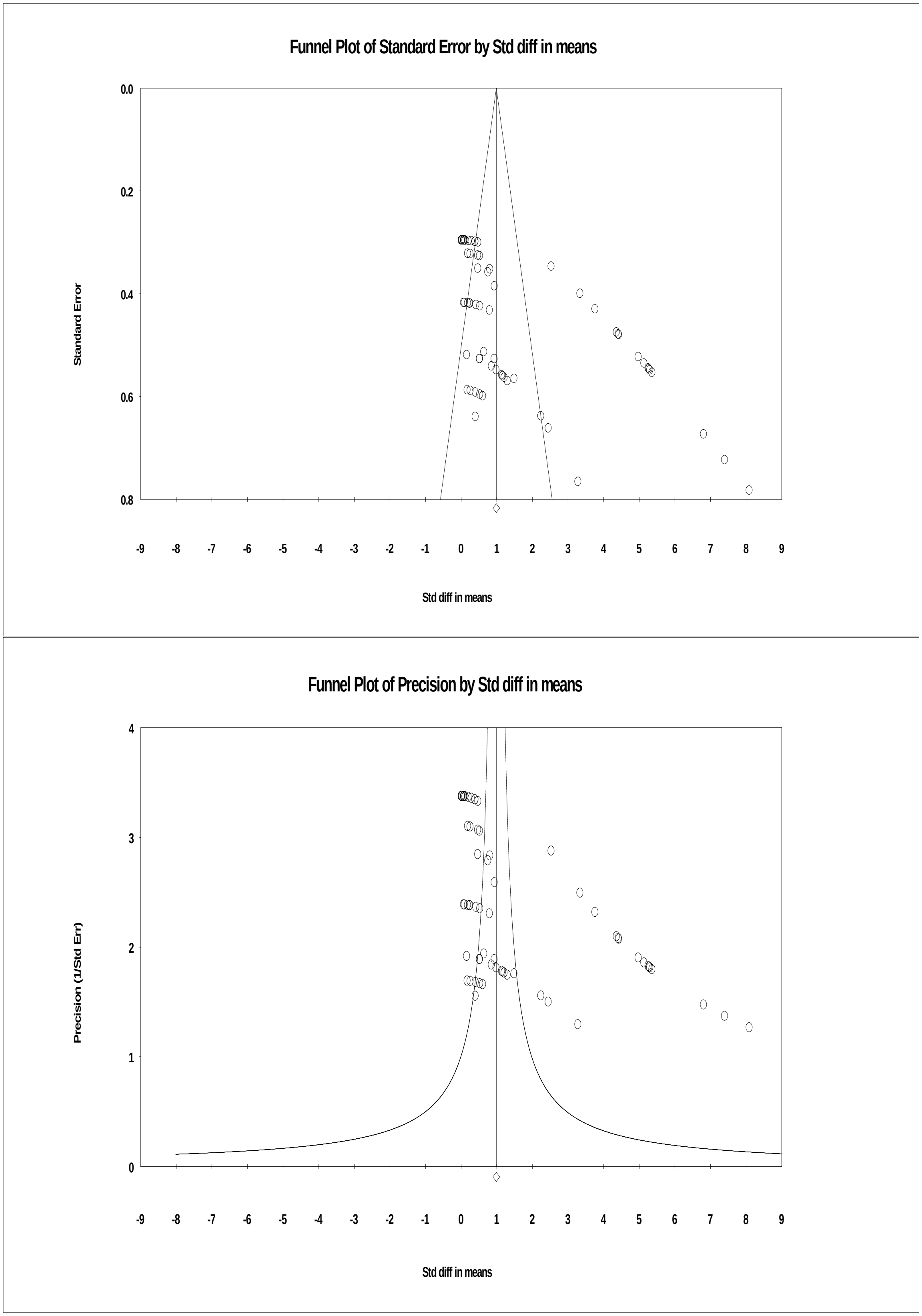


Figure 7. Funnel Plot of all 67 outcomes – fixed effect model

Forest plot of outcome measures by standard difference (Std diff) in means and 95% CI

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Chan (2013)_04	SPE-ATEC	0.248	0.322	0.103	-0.383	0.878	0.770	0.442
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2008)_02	FTS	0.240	0.418	0.175	-0.579	1.060	0.575	0.566
Fragala-Pinkham (2008)_03	Half mile (min)	0.793	0.431	0.186	-0.053	1.638	1.837	0.066
Fragala-Pinkham (2008)_04	Hip abductors (kg)	0.177	0.417	0.174	-0.641	0.996	0.425	0.671
Fragala-Pinkham (2008)_05	Knee extensors (kg)	0.089	0.417	0.174	-0.728	0.906	0.214	0.830
Fragala-Pinkham (2008)_06	Knee flexors (kg)	0.226	0.418	0.175	-0.593	1.045	0.541	0.589
Fragala-Pinkham (2008)_07	Ankle plantarflexors (kg)	0.065	0.417	0.174	-0.752	0.882	0.156	0.876
Fragala-Pinkham (2008)_08	Modified curl-up (count)	0.412	0.421	0.177	-0.412	1.237	0.980	0.327
Fragala-Pinkham (2011)_01	Swimming Classification Scale (levels)	0.602	0.598	0.358	-0.571	1.774	1.006	0.315
Fragala-Pinkham (2011)_02	Mobility skills (scaled scores)	0.166	0.587	0.344	-0.983	1.316	0.284	0.777
Fragala-Pinkham (2011)_03	1/2 Mile	0.396	0.591	0.349	-0.762	1.555	0.671	0.502
Fragala-Pinkham (2011)_04	IPU	0.519	0.595	0.354	-0.647	1.686	0.873	0.383
Fragala-Pinkham (2011)_05	Modified curl-ups (repetitions)	0.252	0.588	0.345	-0.900	1.404	0.428	0.669
Kern (2011)_00	CARS	0.390	0.298	0.089	-0.194	0.974	1.308	0.191
Kern (2011)_01	TPCIS-Expressiveness	0.106	0.295	0.087	-0.472	0.685	0.360	0.718
Kern (2011)_02	TPCIS-Responsiveness	0.000	0.295	0.087	-0.579	0.579	0.000	1.000
Kern (2011)_03	TPCIS-Positive Regard	0.080	0.295	0.087	-0.499	0.658	0.270	0.787
Kern (2011)_04	TPCIS-Negative Regard	0.080	0.295	0.087	-0.499	0.659	0.271	0.786
Kern (2011)_05	TPCIS-Mood and Tone	0.197	0.296	0.088	-0.383	0.777	0.666	0.505
Kern (2011)_06	TPCIS-Empathy	0.063	0.295	0.087	-0.516	0.642	0.214	0.831
Kern (2011)_07	SP-Auditory high-threshold	0.467	0.299	0.089	-0.119	1.053	1.561	0.118
Kern (2011)_08	SP-Auditory low-threshold	0.011	0.295	0.087	-0.567	0.590	0.038	0.969
Kern (2011)_09	SP-Visual high-threshold	0.027	0.295	0.087	-0.552	0.605	0.091	0.928
Kern (2011)_10	SP-Visual low-threshold	0.110	0.295	0.087	-0.469	0.689	0.371	0.711
Kern (2011)_11	SP-Touch high-threshold	0.008	0.295	0.087	-0.571	0.586	0.026	0.980
Kern (2011)_12	SP-Touch low-threshold	0.070	0.295	0.087	-0.509	0.649	0.238	0.812
Kern (2011)_13	SP-Vestibular high-threshold	0.384	0.298	0.089	-0.200	0.968	1.289	0.197
Kern (2011)_14	SP-Vestibular low-threshold	0.283	0.297	0.088	-0.299	0.864	0.953	0.341
Koening (2012)_01	ABC-Community-maladaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_01	HAAR-IV	2.235	0.637	0.406	0.986	3.484	3.507	0.000
Pan (2010)_03	HAAR-II	2.445	0.661	0.437	1.150	3.741	3.700	0.000
Pan (2010)_04	HAAR-III	1.483	0.565	0.319	0.376	2.589	2.626	0.009
Pan (2010)_05	HAAR-V	3.276	0.765	0.585	1.776	4.775	4.282	0.000
Pan (2010)_06	SC-SSBS-2	0.633	0.512	0.263	-0.371	1.637	1.236	0.217
Pan (2010)_07	AnB-SSBS-2	0.926	0.526	0.277	-0.106	1.957	1.759	0.079
Pan (2011)_01	HAAR-IV	1.205	0.562	0.316	0.103	2.307	2.142	0.032
Pan (2011)_04	HAAR-III	0.978	0.547	0.300	-0.095	2.051	1.786	0.074
Pan (2011)_05	HAAR-V	1.293	0.569	0.324	0.178	2.408	2.273	0.023
Pan (2011)_06	Physical fitness-SnR	0.152	0.518	0.269	-0.863	1.168	0.294	0.769
Pan (2011)_07	Physical fitness-Percent body fat	0.513	0.526	0.277	-0.518	1.544	0.975	0.330
Pan (2011)_08	Physical fitness-Curl-ups (30s)	1.163	0.559	0.313	0.067	2.259	2.079	0.038
Pan (2011)_09	Physical fitness-Curl-ups (60s)	1.133	0.557	0.311	0.041	2.226	2.033	0.042
Pan (2011)_10	Physical fitness-16-m PACER	0.511	0.526	0.277	-0.520	1.542	0.972	0.331
Pan (2011)_11	Physical fitness-BMI	0.848	0.540	0.292	-0.211	1.907	1.570	0.116
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Wuang (2010)_01	BOTMP-Running speed and agility	4.422	0.479	0.230	3.483	5.361	9.228	0.000
Wuang (2010)_02	BOTMP-Balance	3.333	0.399	0.159	2.551	4.115	8.352	0.000
Wuang (2010)_03	BOTMP-Bilateral coordination	4.970	0.522	0.273	3.947	5.994	9.521	0.000
Wuang (2010)_04	BOTMP-Strength	5.129	0.535	0.286	4.081	6.177	9.593	0.000
Wuang (2010)_05	BOTMP-Upper-limb coordination	4.357	0.474	0.225	3.428	5.286	9.188	0.000
Wuang (2010)_06	BOTMP-Response speed	6.804	0.673	0.452	5.485	8.122	10.115	0.000
Wuang (2010)_07	BOTMP-Visual-motor control	8.086	0.782	0.611	6.553	9.618	10.340	0.000
Wuang (2010)_08	BOTMP-Upper-limb speed and dexterity	7.394	0.723	0.522	5.978	8.810	10.231	0.000
Wuang (2010)_09	TSIF-Postural-ocular movement	5.294	0.548	0.300	4.220	6.367	9.662	0.000
Wuang (2010)_10	TSIF-Bilateral integration	5.245	0.544	0.296	4.179	6.312	9.642	0.000
Wuang (2010)_11	TSIF-Sensory discrimination	5.355	0.553	0.306	4.271	6.438	9.686	0.000
Wuang (2010)_12	TSIF-Sensory modulation	5.270	0.546	0.298	4.200	6.340	9.652	0.000
Wuang (2010)_13	TSIF-Sensory searching	2.524	0.346	0.120	1.846	3.202	7.294	0.000
Wuang (2010)_14	TSIF-Attention and activity	4.406	0.478	0.228	3.469	5.343	9.219	0.000
Wuang (2010)_15	TSIF-Emotion and behavior	3.755	0.429	0.184	2.914	4.596	8.750	0.000
		1.535	0.199	0.040	1.144	1.926	7.697	0.000

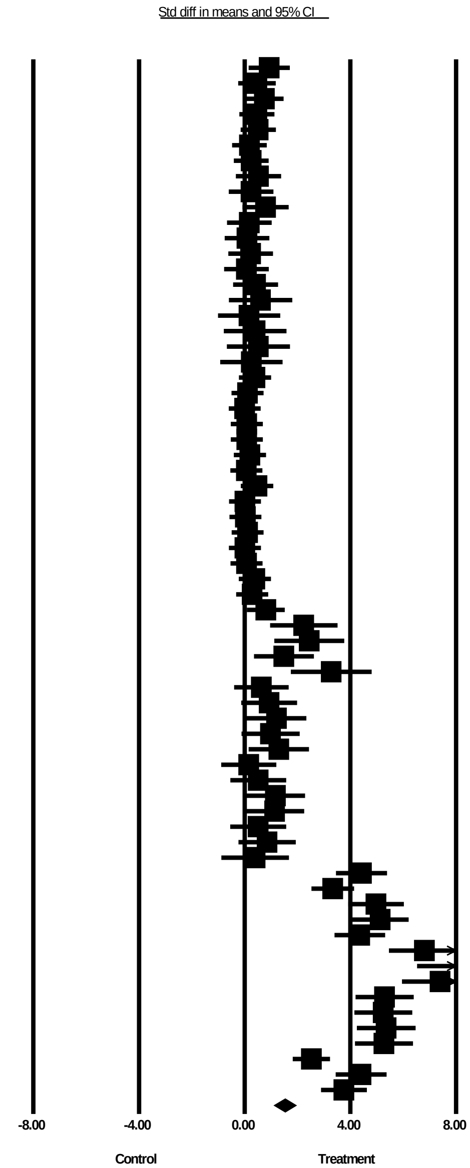


Figure 8. Forest Plot of all 67 outcomes – random effect model

*only the first author and publication year were shown.

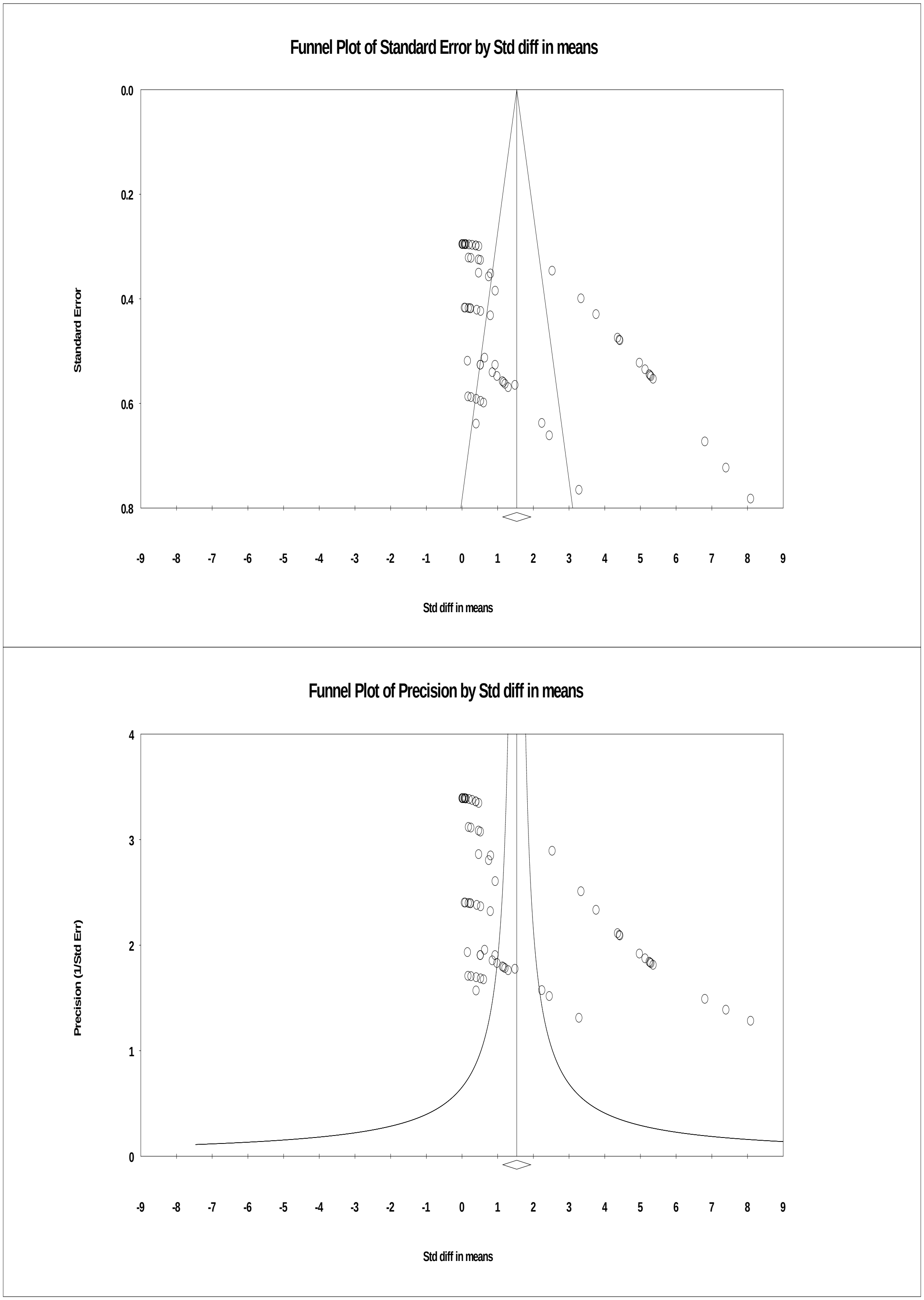


Figure 9. Funnel Plot of all 67 outcomes – random effect model

11 selected studies – lumped outcomes

In the figures 10 and 11, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.987, 95% CI = 0.890, 1.084, p = 0.000) showed the exercise gave a large effect on enhancing performance in all 11 selected studies among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 12 and 13, a random effect model was applied, it showed a statistically insignificant result. The overall mean Cohen's effect size (d = 0.998, 95% CI = -0.019, 2.014, p = 0.054) showed the exercise gave a large effect on enhancing performance in all 11 selected studies among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

For the outcome category of all 11 selected studies, the heterogeneity Q statistic showed a statistical significance (Q = 918.175, df (Q) = 10, p = 0.000), hence, a random effect model was more appropriate in combining the data.

Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)	0.602	0.250	0.062	0.113	1.092	2.410	0.016
Chan (2013)	0.349	0.162	0.026	0.032	0.666	2.159	0.031
Fragala-Pinkham (2008)	0.311	0.149	0.022	0.019	0.602	2.090	0.037
Fragala-Pinkham (2011)	0.385	0.265	0.070	-0.134	0.903	1.454	0.146
Kern (2011)	0.150	0.076	0.006	0.001	0.300	1.969	0.049
Koening (2012)	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)	1.593	0.242	0.059	1.117	2.068	6.570	0.000
Pan (2011)	0.843	0.181	0.033	0.487	1.198	4.646	0.000
Pitetti (2007)	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Wuang (2010)	4.529	0.130	0.017	4.275	4.783	34.933	0.000
	0.987	0.049	0.002	0.890	1.084	19.978	0.000

Statistics for each study

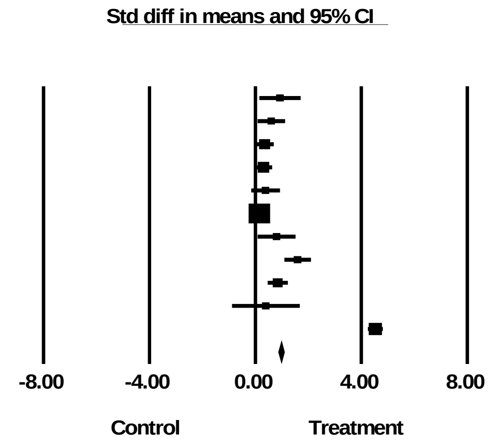


Figure 10. Forest Plot of all 11 selected studies (pooled outcomes) – fixed effect model

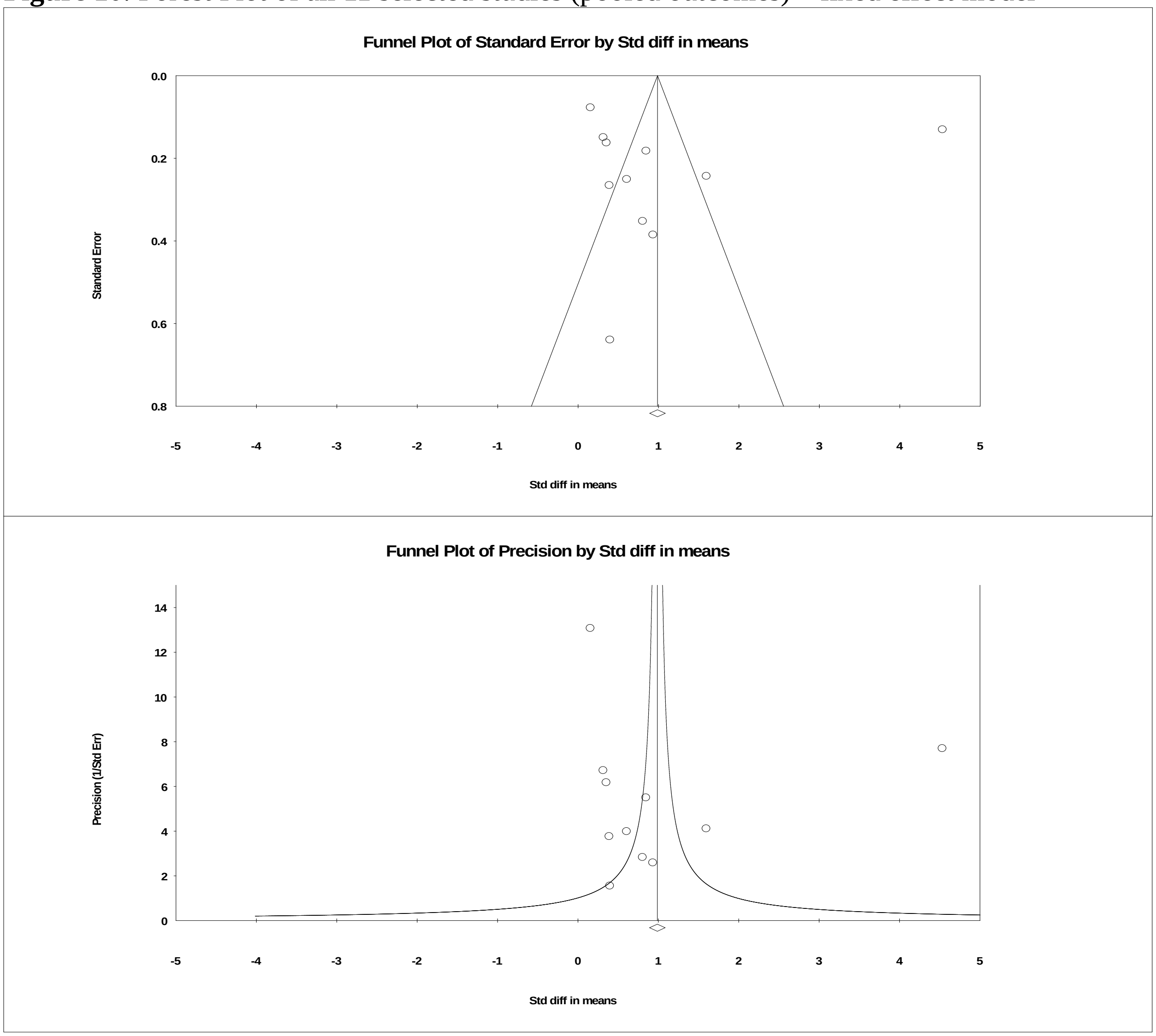


Figure 11. Funnel Plot of all 11 selected studies (pooled outcomes) – fixed effect model

Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)	0.602	0.250	0.062	0.113	1.092	2.410	0.016
Chan (2013)	0.349	0.162	0.026	0.032	0.666	2.159	0.031
Fragala-Pinkham (2008)	0.311	0.149	0.022	0.019	0.602	2.090	0.037
Fragala-Pinkham (2011)	0.385	0.265	0.070	-0.134	0.903	1.454	0.146
Kern (2011)	0.150	0.076	0.006	0.001	0.300	1.969	0.049
Koening (2012)	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)	1.593	0.242	0.059	1.117	2.068	6.570	0.000
Pan (2011)	0.843	0.181	0.033	0.487	1.198	4.646	0.000
Pitetti (2007)	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Wuang (2010)	4.529	0.130	0.017	4.275	4.783	34.933	0.000
	0.998	0.519	0.269	-0.019	2.014	1.924	0.054

(Statistics for each study)

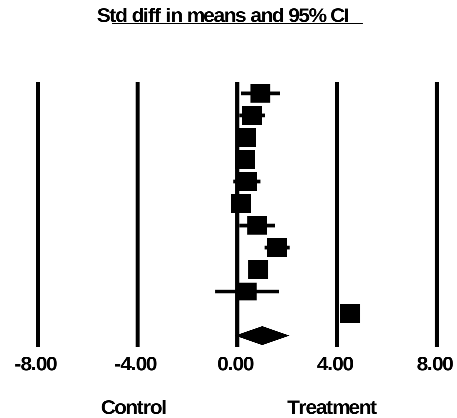
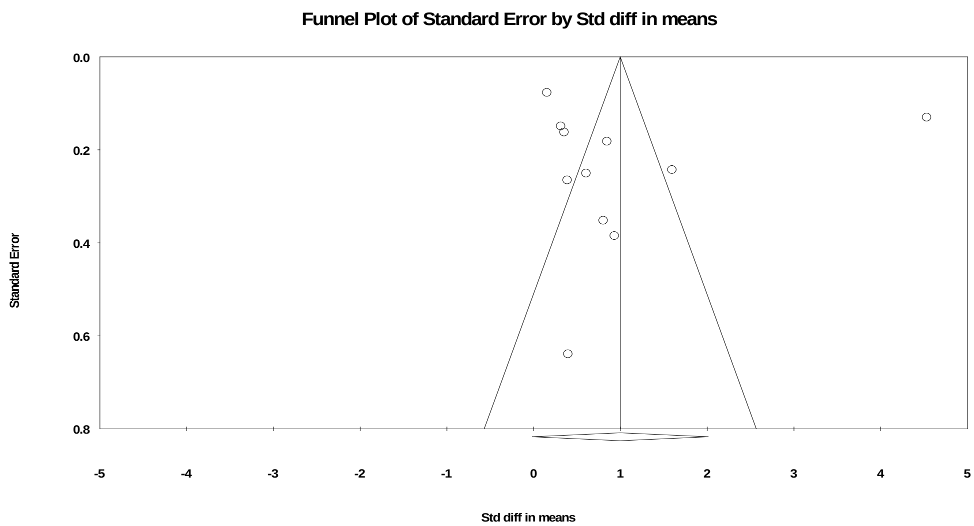


Figure 12. Forest Plot of all 11 selected studies (pooled outcomes) – random effect model



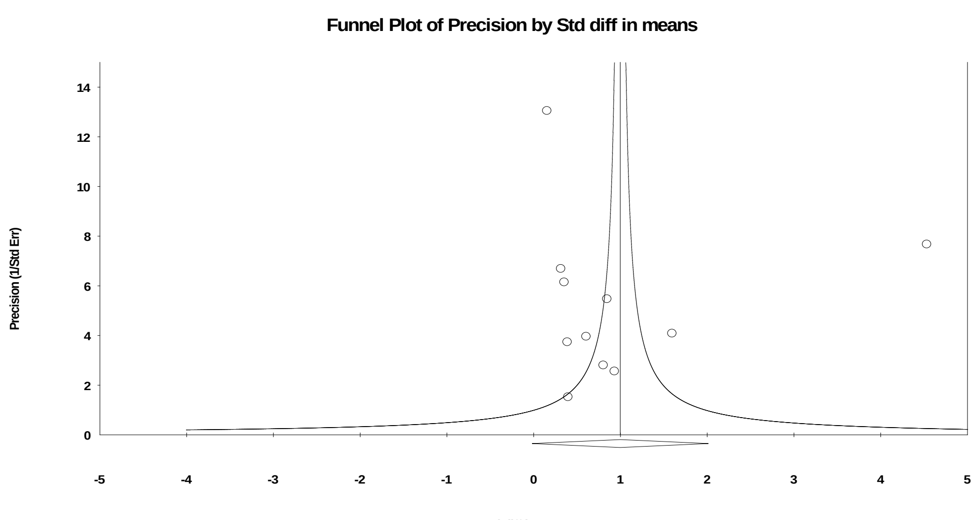


Figure 13. Funnel Plot of all 11 selected studies (pooled outcomes) – random effect model

Exercise Performance & Sport/Skill-related Fitness (EXE)

In the figures 14 and 15, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen’s effect size (d = 1.049, 95% CI = 0.896, 1.202, p = 0.000) showed the exercise gave a large effect on enhancing performance in Exercise Performance and Sport/Skill-related Fitness (EXE) among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 16 and 17, a random effect model was applied, it showed a statistically insignificant result. The overall mean Cohen’s effect size (d = 1.246, 95% CI = -0.086, 2.577, p = 0.067) showed the exercise gave a large effect on enhancing performance in Exercise Performance and Sport/Skill-related Fitness (EXE) among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

For the outcome category of Exercise Performance and Sport/Skill-related Fitness (EXE), the heterogeneity Q statistic showed a statistical significance (Q = 390.503, df (Q) = 7, p = 0.000), hence, a random effect model was more appropriate in combining the data.

Exercise Performance & Sport/Skill-related Fitness (EXE) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2011)_01	SCS & Mobility skills	0.380	0.419	0.175	-0.441	1.201	0.907	0.365
Pan (2010)_01	HAAR	2.224	0.323	0.104	1.592	2.857	6.891	0.000
Pan (2011)_01	HAAR	1.154	0.323	0.104	0.521	1.787	3.574	0.000
Kern (2011)_01	SP	0.168	0.105	0.011	-0.037	0.374	1.607	0.108
Wuang (2010)_02	BOTMP-VU, TSIF-PBSSS	4.529	0.202	0.041	4.133	4.925	22.399	0.000
		1.049	0.078	0.006	0.896	1.202	13.424	0.000

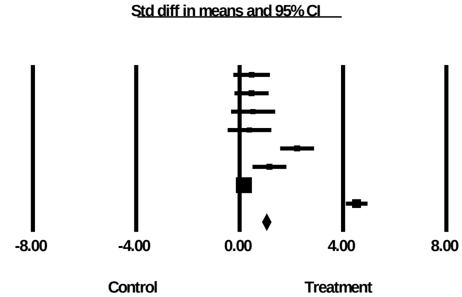


Figure 14. Forest Plot of Exercise Performance & Sport/Skill-related Fitness (EXE) – fixed effect model
*only the first author and publication year were shown

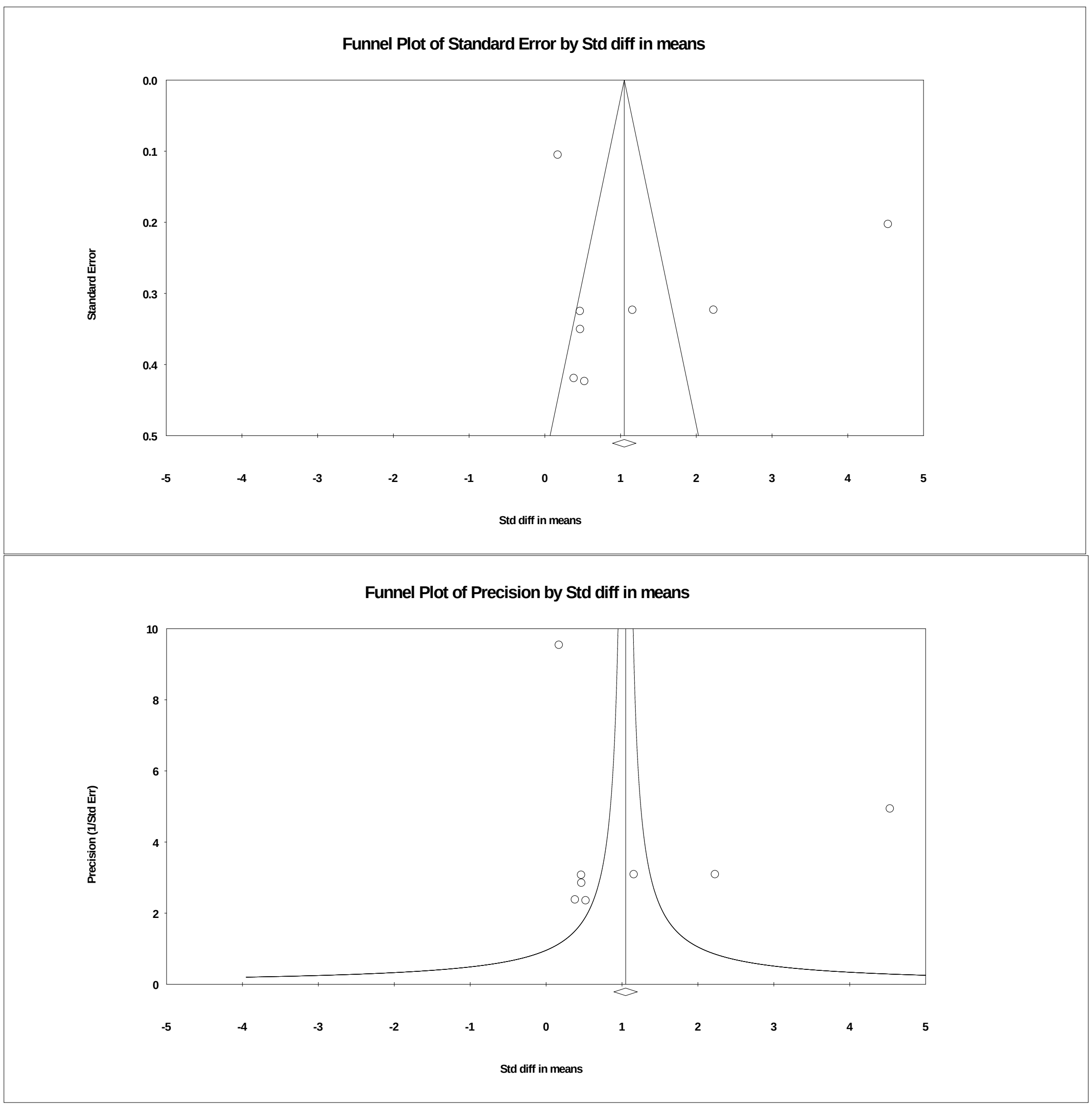


Figure 15. Funnel Plot of Exercise Performance & Sport/Skill-related Fitness (EXE) – fixed effect model

Exercise Performance & Sport/Skill-related Fitness (EXE) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2011)_01	SCS & Mobility skills	0.380	0.419	0.175	-0.441	1.201	0.907	0.365
Pan (2010)_01	HAAR	2.224	0.323	0.104	1.592	2.857	6.891	0.000
Pan (2011)_01	HAAR	1.154	0.323	0.104	0.521	1.787	3.574	0.000
Kern (2011)_01	SP	0.168	0.105	0.011	-0.037	0.374	1.607	0.108
Wuang (2010)_02	BOTMP-VU, TSIF-PBSSS	4.529	0.202	0.041	4.133	4.925	22.399	0.000
		1.246	0.679	0.462	-0.086	2.577	1.834	0.067

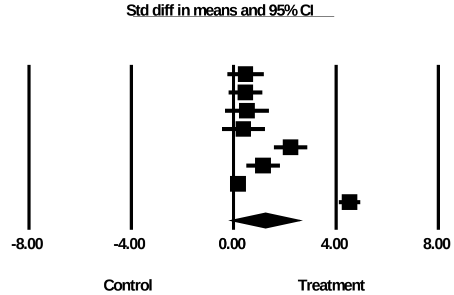


Figure 16. Forest Plot of Exercise Performance & Sport/Skill-related Fitness (EXE) – random effect model

*only the first author and publication year were shown

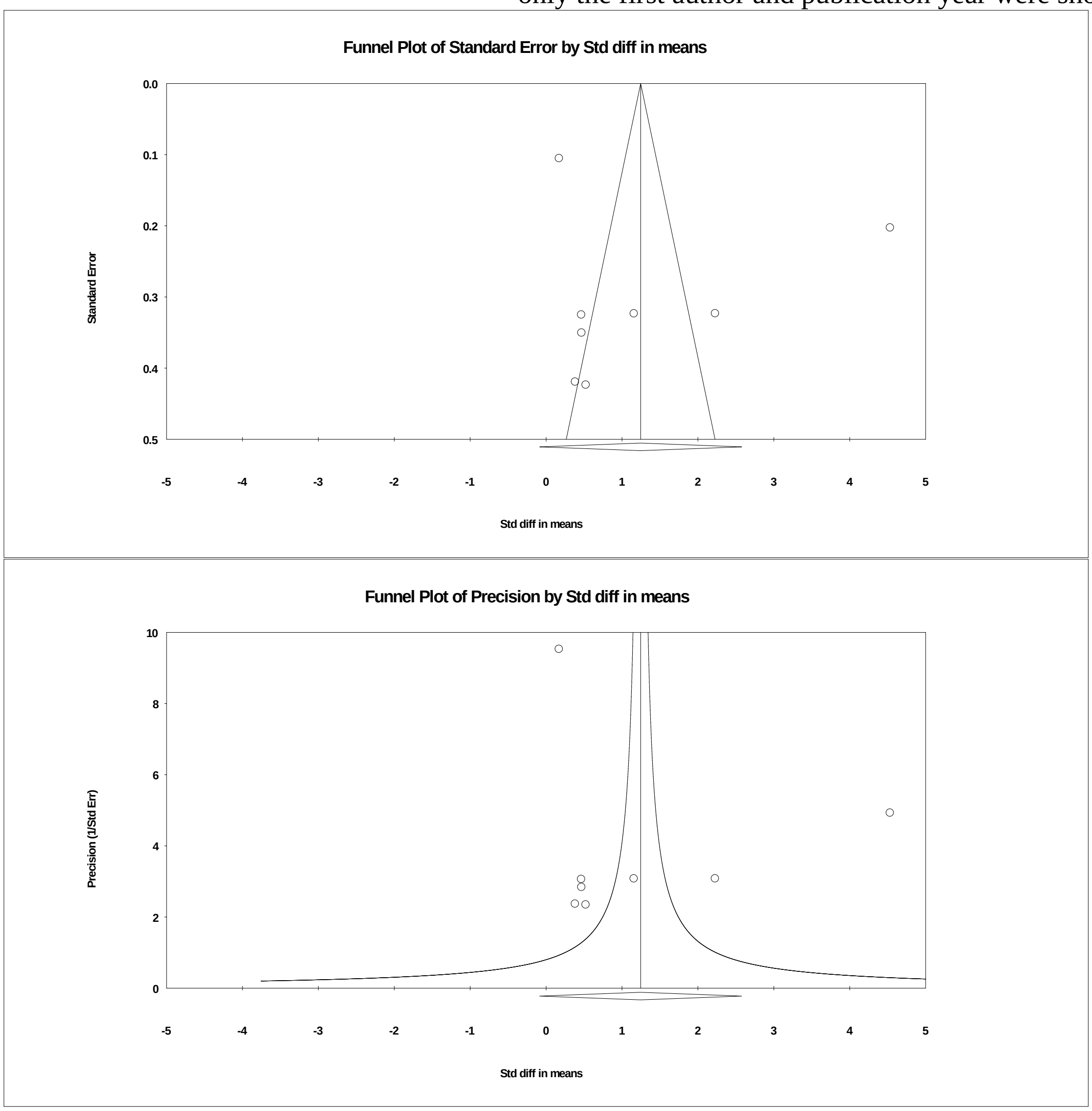


Figure 17. Funnel Plot of Exercise Performance & Sport/Skill-related Fitness (EXE) – random effect model

Physiological & Biometric Indicator (PHY)

In the figures 18 and 19, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 1.469, 95% CI = 1.275, 1.662, p = 0.000) showed the exercise gave a large effect on enhancing performance in Physiological and Biometric Indicator (PHY) among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 20 and 21, a random effect model was applied, it showed a statistically insignificant result. The overall mean Cohen's effect size (d = 1.174, 95% CI = -0.558, 2.905, p = 0.184) showed the exercise gave a large effect on enhancing performance in Physiological and Biometric Indicator (PHY) among children with ASD. The funnel plots showed a left-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

For the outcome category of Physiological and Biometric Indicator (PHY), the heterogeneity Q statistic showed a statistical significance (Q = 354.662, df (Q) = 5, p = 0.000), hence, a random effect model was more appropriate in combining the data.

Physiological & Biometric Indicator (PHY) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Fragala-Pinkham (2008)_02	FTS, 1/2 Mile, Joint ROM, & Modified Curl-up	0.281	0.159	0.025	-0.030	0.592	1.771	0.077
Fragala-Pinkham (2011)_02	1/2 Mile, IPU, & Modified Curl-up	0.388	0.341	0.117	-0.281	1.057	1.137	0.256
Pan (2011)_02	SnR, Percent Body Fat, Curl-up(30s), Curl-up(60s), & PACER	0.670	0.240	0.058	0.200	1.140	2.791	0.005
Wuang (2010)_01	BOTMP-RBBSUR	4.716	0.199	0.040	4.326	5.106	23.691	0.000
		1.469	0.099	0.010	1.275	1.662	14.889	0.000

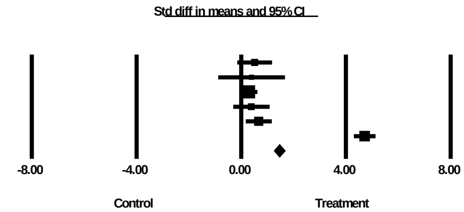


Figure 18. Forest Plot of Physiological & Biometric Indicator (PHY) – fixed effect model
*only the first author and publication year were shown

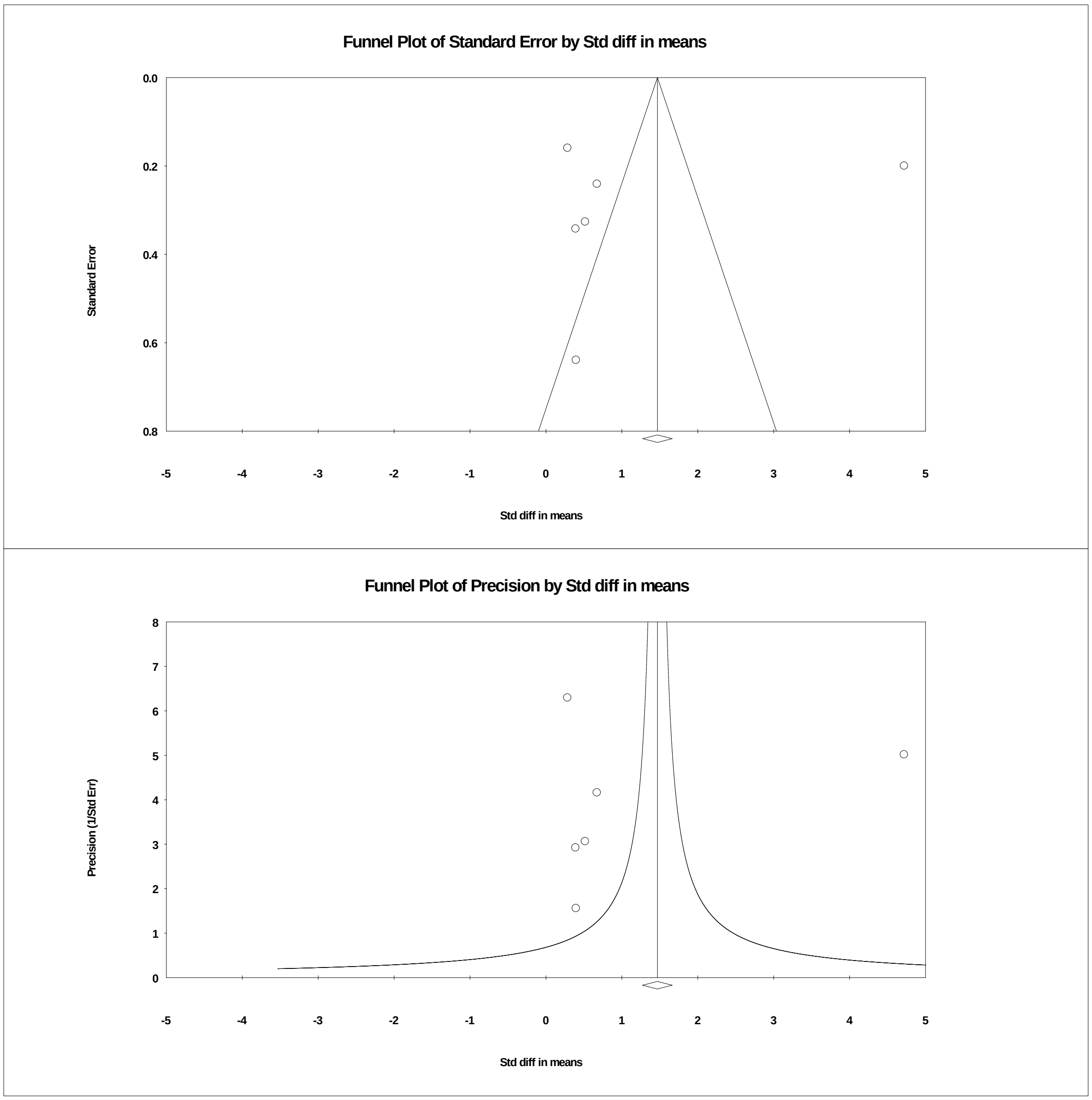


Figure 19. Funnel Plot of Physiological & Biometric Indicator (PHY) – fixed effect model

Physiological & Biometric Indicator (PHY) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Fragala-Pinkham (2008)_02	FTS, 1/2 Mile, Joint ROM, & Modified Curl-up	0.281	0.159	0.025	-0.030	0.592	1.771	0.077
Fragala-Pinkham (2011)_02	1/2 Mile, IPU, & Modified Curl-up	0.388	0.341	0.117	-0.281	1.057	1.137	0.256
Pan (2011)_02	SnR, Percent Body Fat, Curl-up(30s), Curl-up(60s), & PACER	0.670	0.240	0.058	0.200	1.140	2.791	0.005
Wuang (2010)_01	BOTMP-RBBSUR	4.716	0.199	0.040	4.326	5.106	23.691	0.000
		1.174	0.883	0.780	-0.558	2.905	1.328	0.184

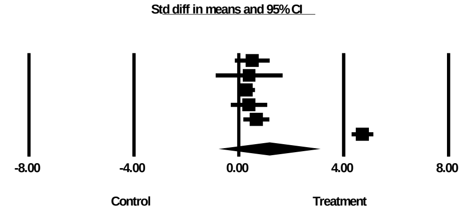


Figure 20. Forest Plot of Physiological & Biometric Indicator (PHY) – random effect model
*only the first author and publication year were shown

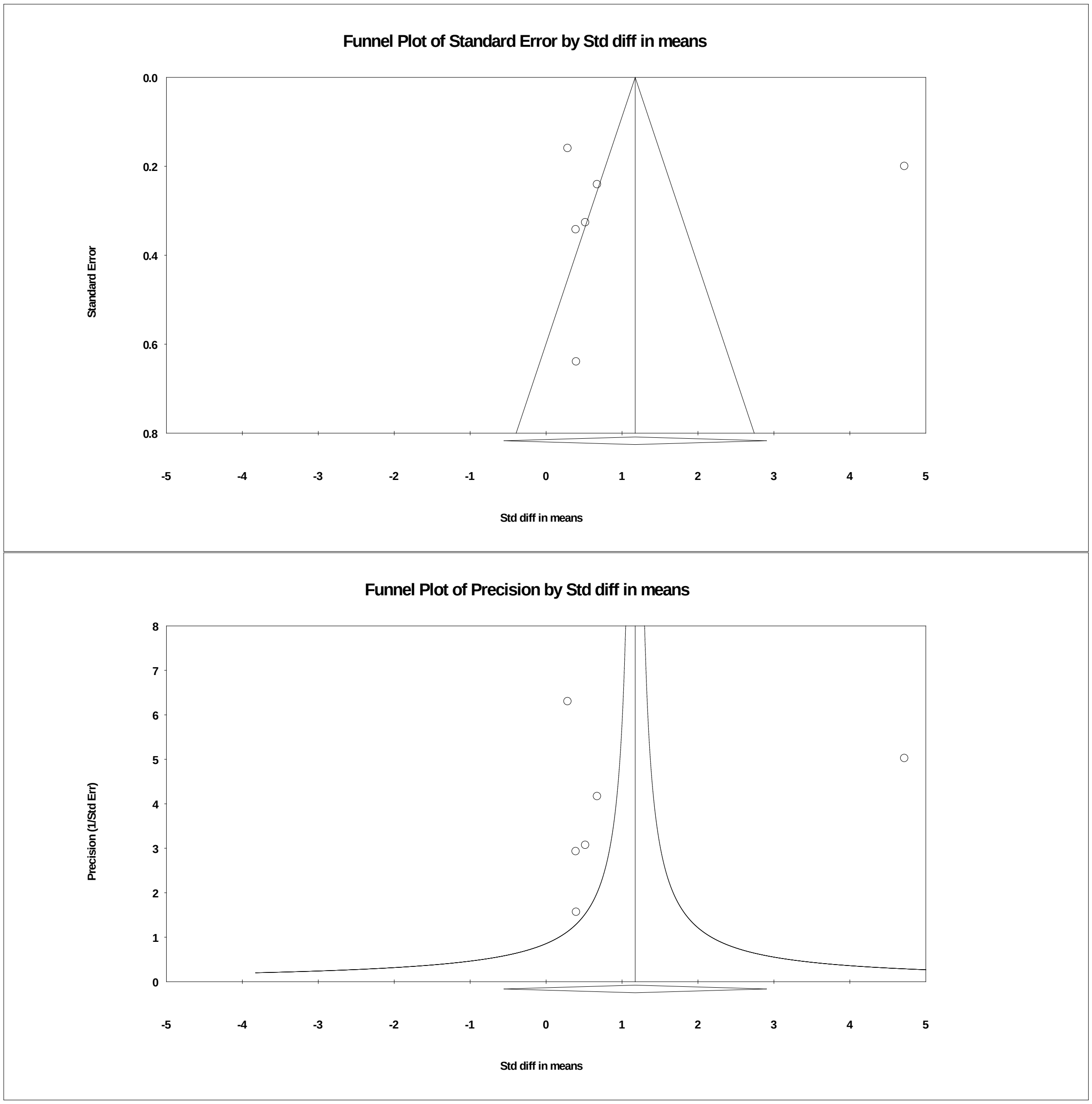


Figure 21. Funnel Plot of Physiological & Biometric Indicator (PHY) – random effect model

Social Cognition & Psychological Well-being (SOC)

In the figures 22 and 23, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.606, 95% CI = 0.426, 0.787, p = 0.000) showed the exercise gave a medium to large effect on enhancing performance in Social Cognition and Psychological Well-being (SOC) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

In the figures 24 and 25, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 1.076, 95% CI = 0.065, 2.087, p = 0.037) showed the exercise gave a large effect on enhancing performance in Social Cognition and Psychological Well-being (SOC) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

For the outcome category of Social Cognition and Psychological Well-being (SOC), the heterogeneity Q statistic showed a statistical significance (Q = 137.659, df (Q) = 6, p = 0.000), hence, a random effect model was more appropriate in combining the data.

Social Cognition & Psychological Well-being (SOC) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Koening (2012)_01	ABC-Community-adaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_02	SSBS-2	0.775	0.367	0.135	0.056	1.495	2.113	0.035
Kern (2011)_02	TPCIS	0.088	0.121	0.015	-0.149	0.324	0.726	0.468
Wuang (2010)_03	TSIF-AE	4.046	0.319	0.102	3.420	4.672	12.670	0.000
		0.606	0.092	0.008	0.426	0.787	6.599	0.000

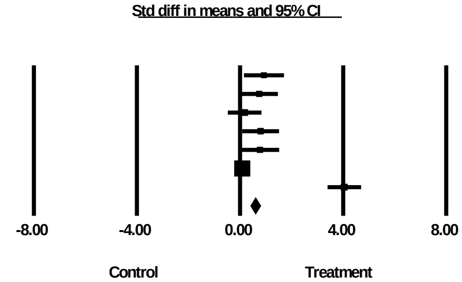


Figure 22. Forest Plot of Social Cognition & Psychological Well-being (SOC) – fixed effect model
*only the first author and publication year were shown

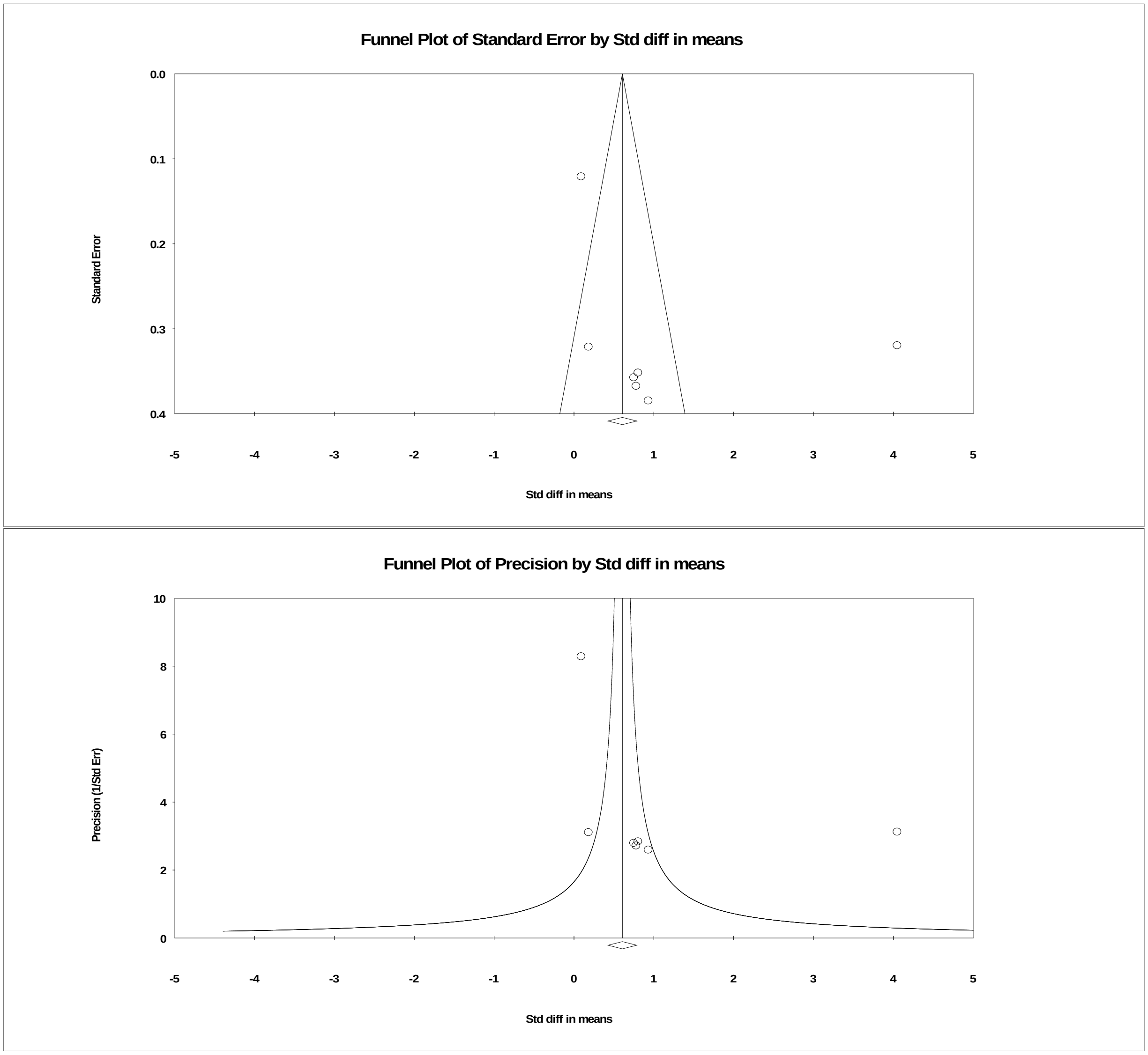


Figure 23. Funnel Plot of Social Cognition & Psychological Well-being (SOC) – fixed effect model

Social Cognition & Psychological Well-being (SOC) (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Koening (2012)_01	ABC-Community-adaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_02	SSBS-2	0.775	0.367	0.135	0.056	1.495	2.113	0.035
Kern (2011)_02	TPCIS	0.088	0.121	0.015	-0.149	0.324	0.726	0.468
Wuang (2010)_03	TSIF-AE	4.046	0.319	0.102	3.420	4.672	12.670	0.000
		1.076	0.516	0.266	0.065	2.087	2.085	0.037

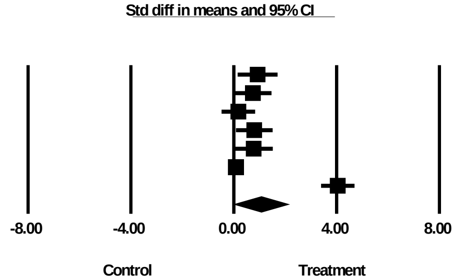
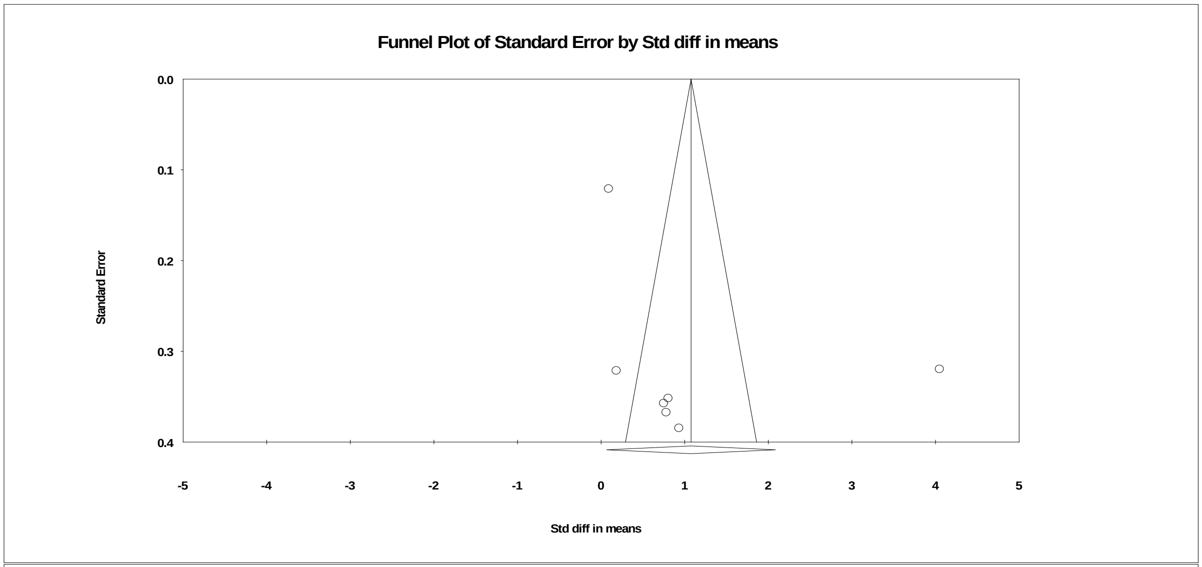


Figure 24. Forest Plot of Social Cognition & Psychological Well-being (SOC) – random effect model
*only the first author and publication year were shown



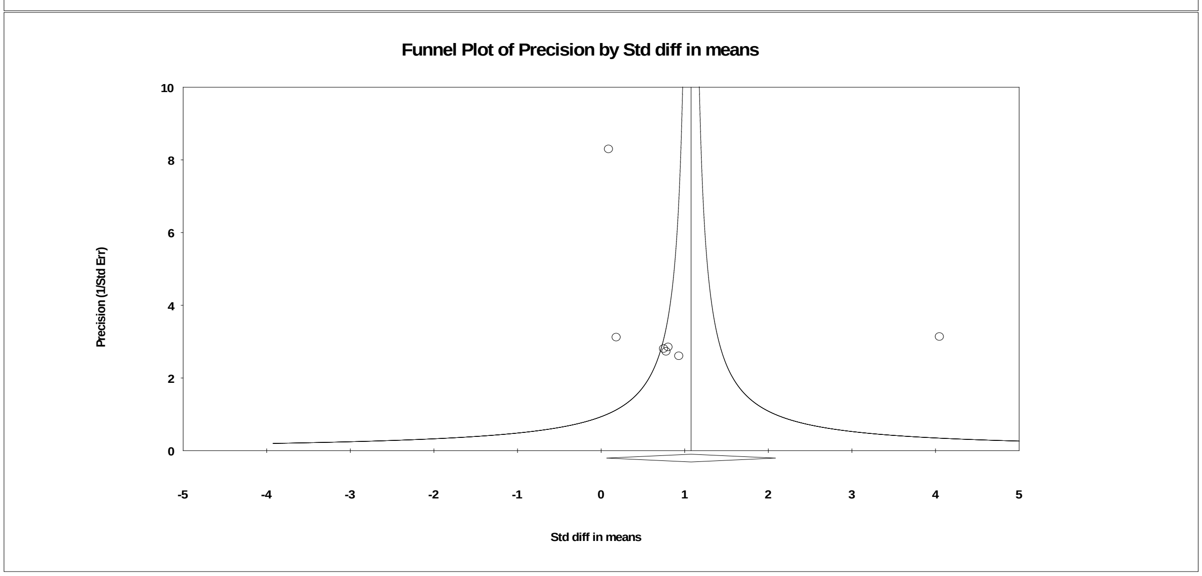


Figure 25. Funnel Plot of Social Cognition & Psychological Well-being (SOC) – random effect model

10 selected studies – lumped outcomes

In the figures 26 and 27, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.385, 95% CI = 0.280, 0.490, p = 0.000) showed the exercise gave a small to medium effect on enhancing performance in 10 selected studies (adjusted) among children with ASD. The funnel plots showed a right-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 28 and 29, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.612, 95% CI = 0.328, 0.895, p = 0.000) showed the exercise gave a medium to large effect on enhancing performance in 10 selected studies (adjusted) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

For the outcome category of 10 selected studies (adjusted), the heterogeneity Q statistic showed a statistical significance (Q = 45.064, df (Q) = 9, p = 0.000), hence, a random effect model was more appropriate in combining the data.

10 selected studies – lumped outcomes (continue)

Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)	0.602	0.250	0.062	0.113	1.092	2.410	0.016
Chan (2013)	0.349	0.162	0.026	0.032	0.666	2.159	0.031
Fragala-Pinkham (2008)	0.311	0.149	0.022	0.019	0.602	2.090	0.037
Fragala-Pinkham (2011)	0.385	0.265	0.070	-0.134	0.903	1.454	0.146
Kern (2011)	0.150	0.076	0.006	0.001	0.300	1.969	0.049
Koening (2012)	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)	1.593	0.242	0.059	1.117	2.068	6.570	0.000
Pan (2011)	0.843	0.181	0.033	0.487	1.198	4.646	0.000
Pitetti (2007)	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
	0.385	0.053	0.003	0.280	0.490	7.205	0.000

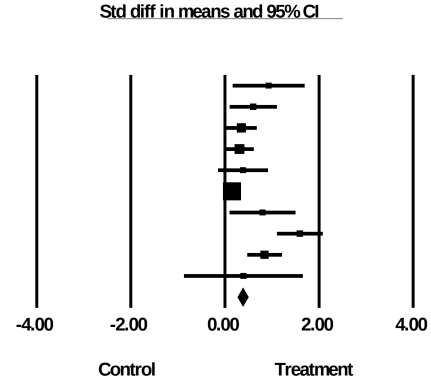


Figure 26. Forest Plot of 10 selected studies (pooled outcomes) – fixed effect model
*only the first author and publication year were shown

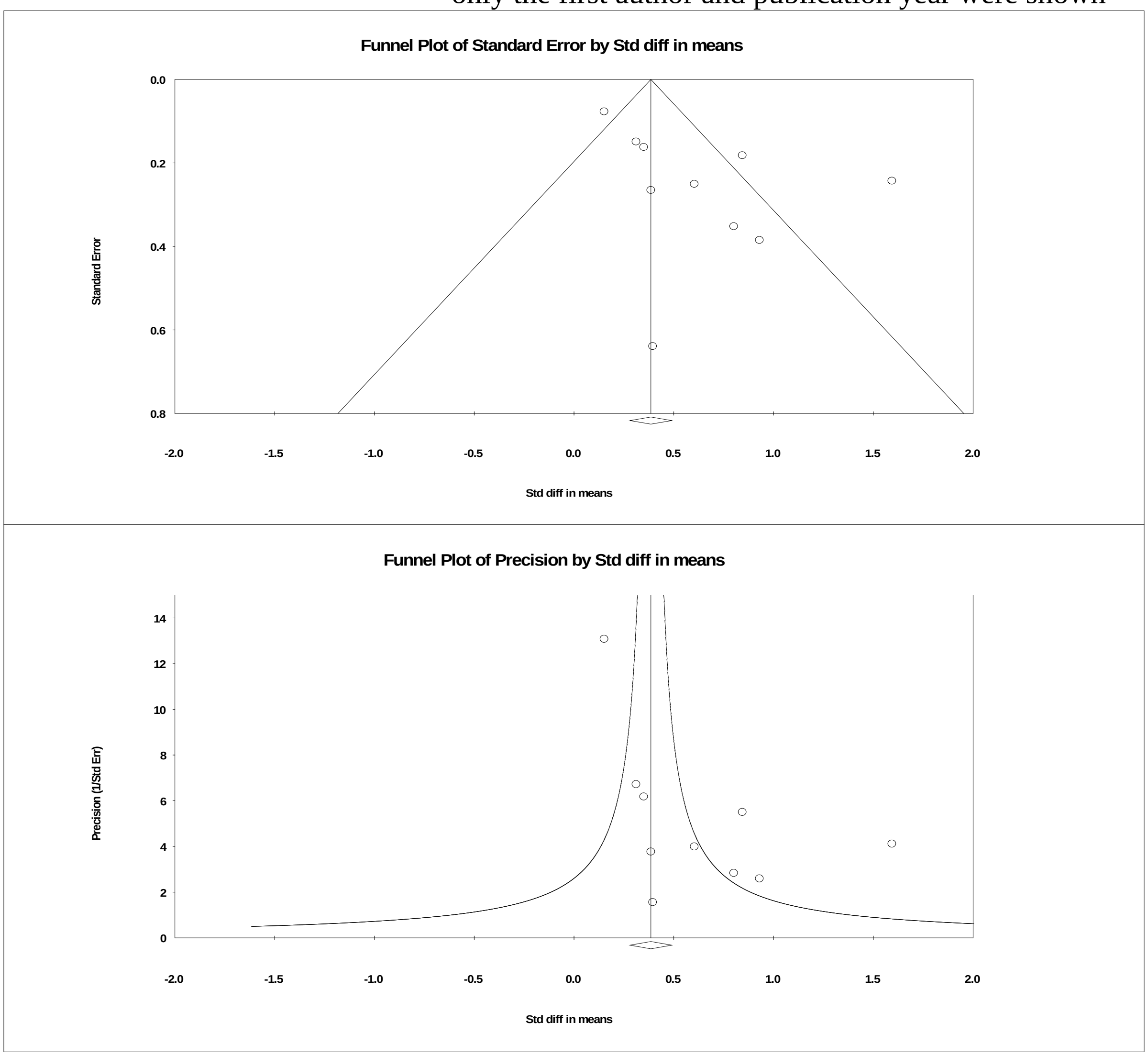


Figure 27. Funnel Plot of 10 selected studies (pooled outcomes) – fixed effect model

10 selected studies – lumped outcomes (continue)

Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Statistics for each study						
Bahrami (2012)	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)	0.602	0.250	0.062	0.113	1.092	2.410	0.016
Chan (2013)	0.349	0.162	0.026	0.032	0.666	2.159	0.031
Fragala-Pinkham (2008)	0.311	0.149	0.022	0.019	0.602	2.090	0.037
Fragala-Pinkham (2011)	0.385	0.265	0.070	-0.134	0.903	1.454	0.146
Kern (2011)	0.150	0.076	0.006	0.001	0.300	1.969	0.049
Koening (2012)	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)	1.593	0.242	0.059	1.117	2.068	6.570	0.000
Pan (2011)	0.843	0.181	0.033	0.487	1.198	4.646	0.000
Pitetti (2007)	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
	0.612	0.145	0.021	0.328	0.895	4.225	0.000

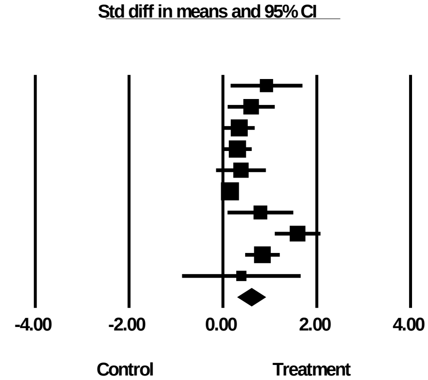


Figure 28. Forest Plot of 10 selected studies (pooled outcomes) – random effect model

*only the first author and publication year were shown

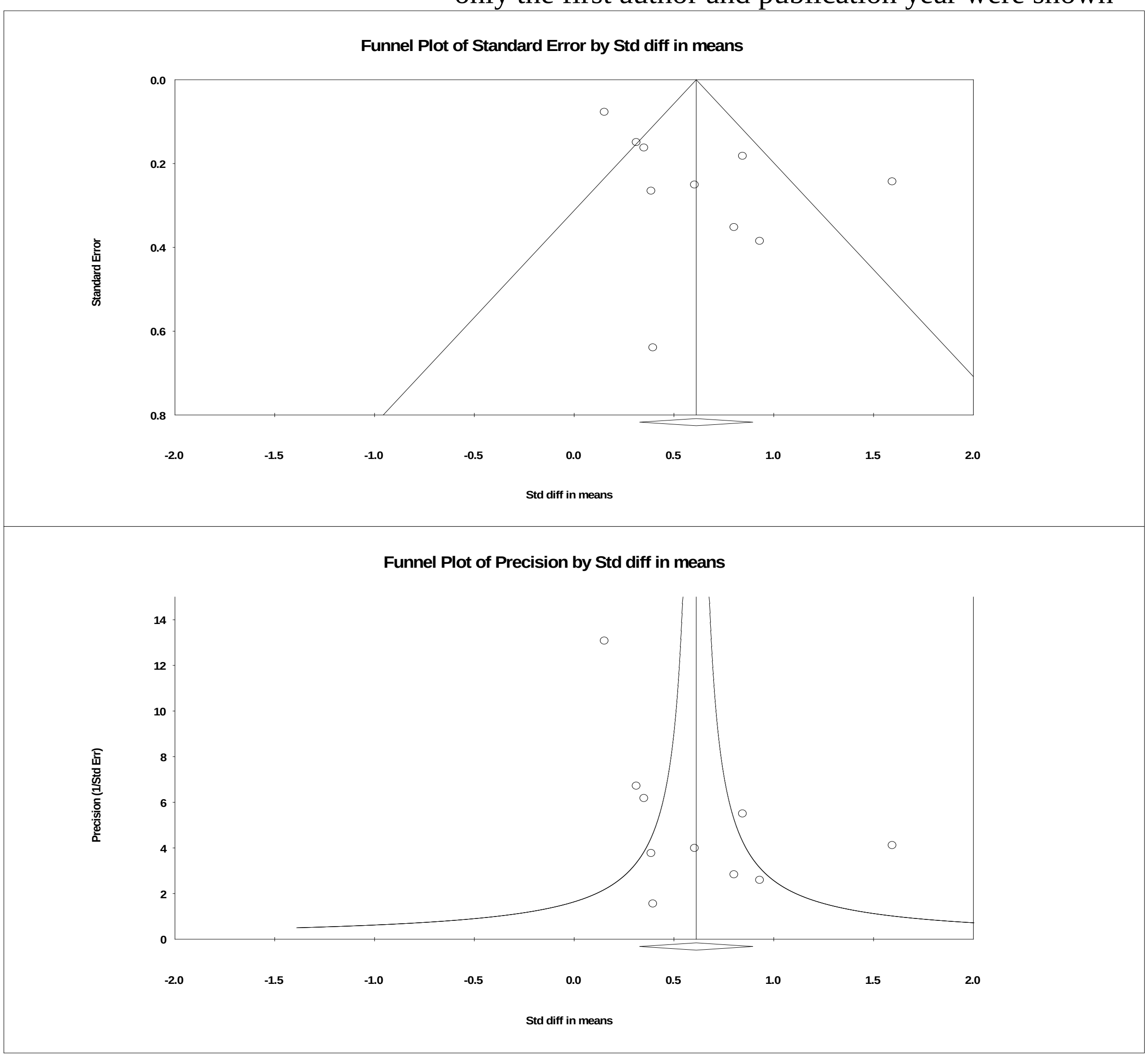


Figure 29. Funnel Plot of 10 selected studies (pooled outcomes) – random effect model

EXE adjusted

In the figures 30 and 31, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.438, 95% CI = 0.272, 0.604, p = 0.000) showed the exercise gave a small to medium effect on enhancing performance in Exercise Performance and Sport/Skill-related Fitness (adjusted EXE) among children with ASD. The funnel plots showed a right-skewed asymmetrical shape, which indicated a publication bias was observed among the included studies.

In the figures 32 and 33, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size (d = 0.763, 95% CI = 0.197, 1.328, p = 0.008) showed the exercise gave a medium to large effect on enhancing performance in Exercise Performance and Sport/Skill-related Fitness (adjusted EXE) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

For the outcome category of Exercise Performance and Sport/Skill-related Fitness (adjusted EXE), the heterogeneity Q statistic showed a statistical significance (Q = 42.250, df (Q) = 6, p = 0.000), hence, a random effect model was more appropriate in combining the data.

EXE adjusted (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2011)_01	SCS & Mobility skills	0.380	0.419	0.175	-0.441	1.201	0.907	0.365
Pan (2010)_01	HAAR	2.224	0.323	0.104	1.592	2.857	6.891	0.000
Pan (2011)_01	HAAR	1.154	0.323	0.104	0.521	1.787	3.574	0.000
Kern (2011)_01	SP	0.168	0.105	0.011	-0.037	0.374	1.607	0.108
		0.438	0.085	0.007	0.272	0.604	5.168	0.000

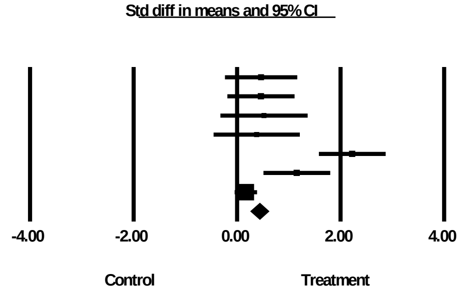


Figure 30. Forest Plot of Exercise Performance & Sport/Skill-related Fitness (adjusted EXE) – fixed effect model

*only the first author and publication year were shown

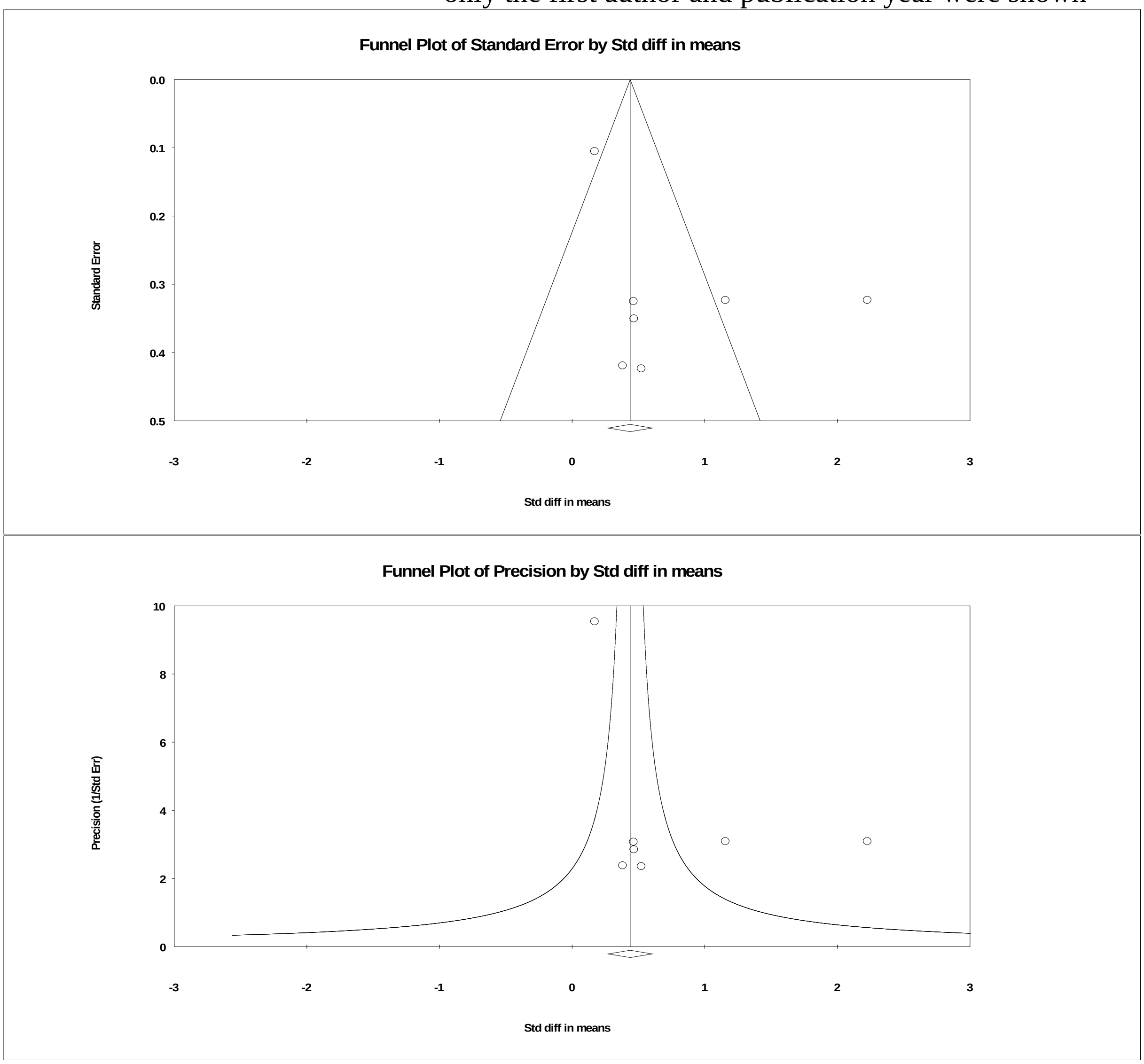


Figure 31. Funnel Plot of Exercise Performance & Sport/Skill-related Fitness (adjusted EXE) – fixed effect model

EXE adjusted (continue)

Study name	Outcome	Statistics for each study						
		Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bass (2009)_01	SP	0.464	0.350	0.122	-0.222	1.150	1.327	0.185
Chan (2013)_01	SEN-ATEC	0.461	0.325	0.105	-0.175	1.097	1.421	0.155
Fragala-Pinkham (2008)_01	M-PEDI	0.520	0.423	0.179	-0.309	1.350	1.230	0.219
Fragala-Pinkham (2011)_01	SCS & Mobility skills	0.380	0.419	0.175	-0.441	1.201	0.907	0.365
Pan (2010)_01	HAAR	2.224	0.323	0.104	1.592	2.857	6.891	0.000
Pan (2011)_01	HAAR	1.154	0.323	0.104	0.521	1.787	3.574	0.000
Kern (2011)_01	SP	0.168	0.105	0.011	-0.037	0.374	1.607	0.108
		0.763	0.288	0.083	0.197	1.328	2.644	0.008

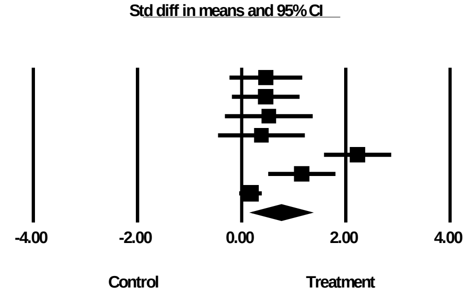


Figure 32. Forest Plot of Exercise Performance & Sport/Skill-related Fitness (adjusted EXE) – random effect model

*only the first author and publication year were shown

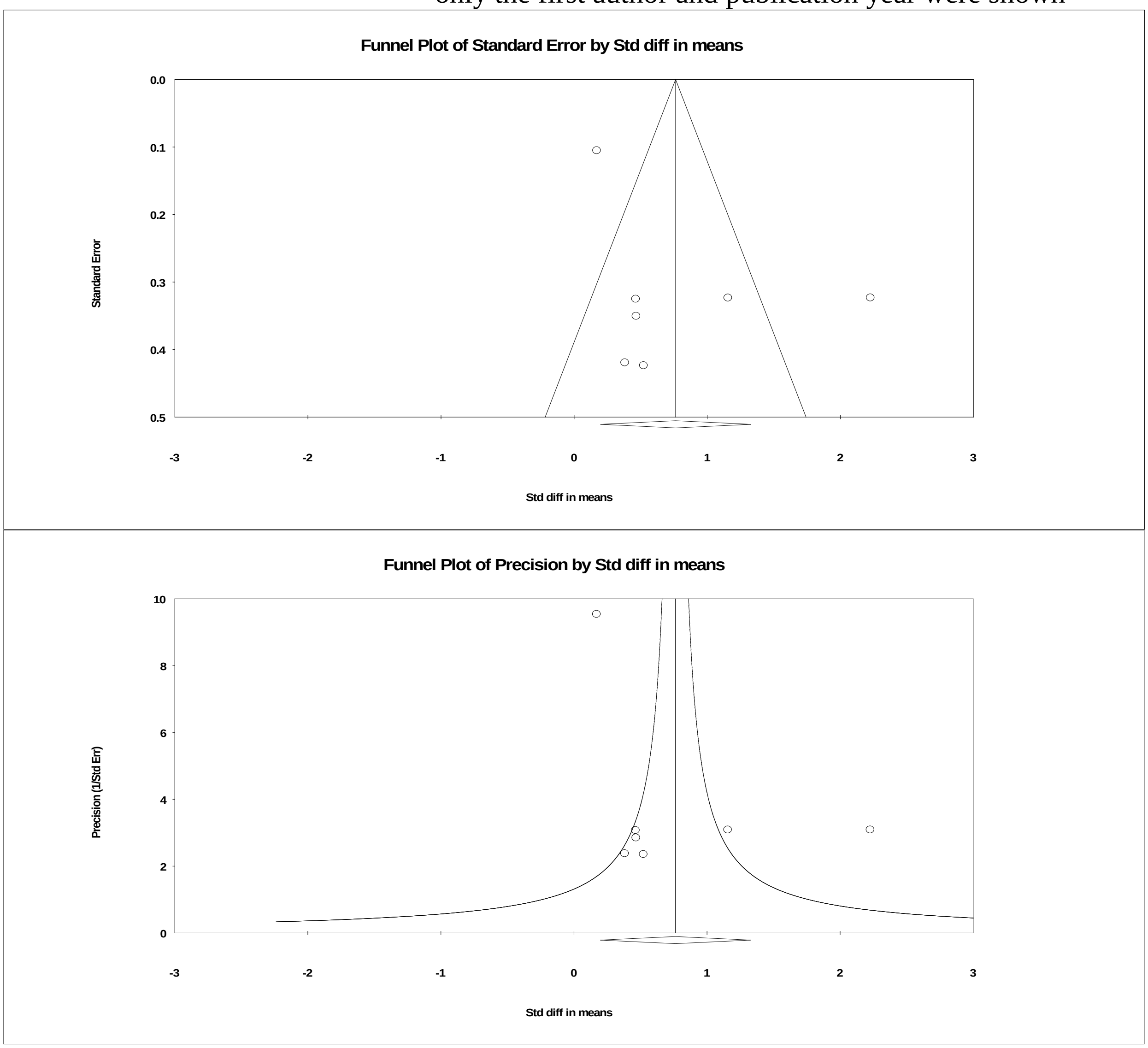


Figure 33. Funnel Plot of Exercise Performance & Sport/Skill-related Fitness (adjusted EXE) – random effect model

PHY adjusted

In the figure 34 and 35, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size ($d = 0.412$, 95% CI = 0.189, 0.634, $p = 0.000$) showed the exercise gave a small to medium effect on enhancing performance in Physiological and Biometric Indicator (adjusted PHY) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

In the figure 36 and 37, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size ($d = 0.412$, 95% CI = 0.189, 0.634, $p = 0.000$) showed the exercise gave a small to medium effect on enhancing performance in Physiological and Biometric Indicator (adjusted PHY) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

For the outcome category of Physiological and Biometric Indicator (adjusted PHY), the heterogeneity Q statistic showed no statistical significance ($Q = 1.939$, $df (Q) = 4$, $p = 0.747$), hence, a fixed effect model was appropriate in combining the data.

PHY adjusted (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Fragala-Pinkham (2008)_02	FTS, 1/2 Mile, Joint ROM, & Modified Curl-up	0.281	0.159	0.025	-0.030	0.592	1.771	0.077
Fragala-Pinkham (2011)_02	1/2 Mile, IPU, & Modified Curl-up	0.388	0.341	0.117	-0.281	1.057	1.137	0.256
Pan (2011)_02	SnR, Percent Body Fat, Curl-up(30s), Curl-up(60s), & PACER	0.670	0.240	0.058	0.200	1.140	2.791	0.005
		0.412	0.114	0.013	0.189	0.634	3.626	0.000

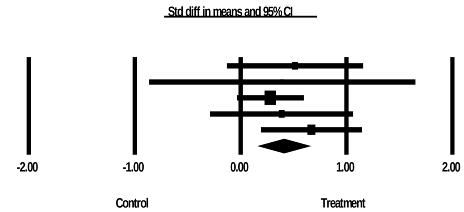


Figure 34. Forest Plot of Physiological & Biometric Indicator (adjusted PHY) – fixed effect model

*only the first author and publication year were shown

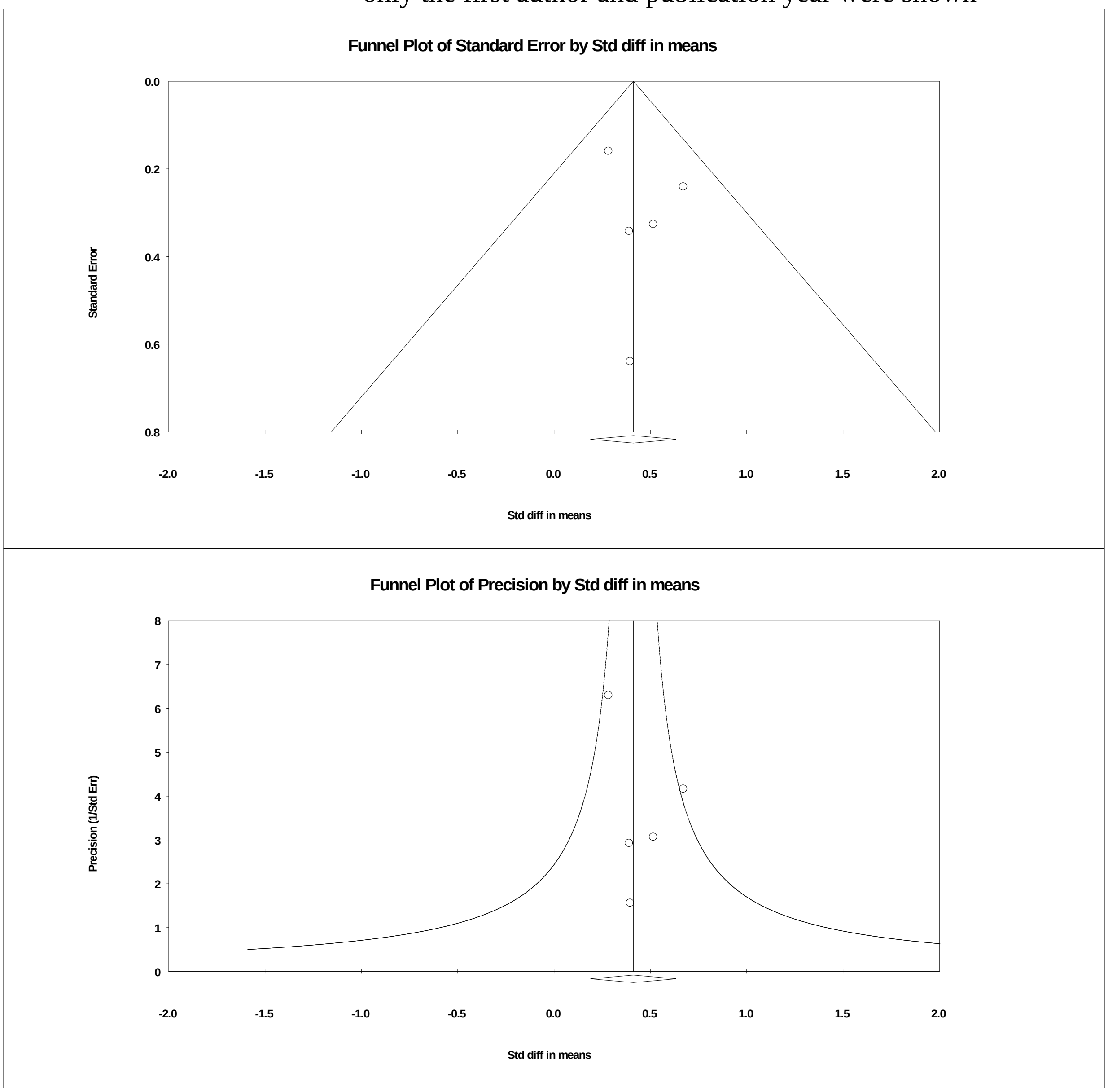


Figure 35. Funnel Plot of Physiological & Biometric Indicator (adjusted PHY) – fixed effect model

PHY adjusted (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Chan (2013)_02	PHY-ATEC	0.514	0.326	0.106	-0.124	1.152	1.578	0.115
Pitetti (2007)_01	BMI	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Fragala-Pinkham (2008)_02	FTS, 1/2 Mile, Joint ROM, & Modified Curl-up	0.281	0.159	0.025	-0.030	0.592	1.771	0.077
Fragala-Pinkham (2011)_02	1/2 Mile, IPU, & Modified Curl-up	0.388	0.341	0.117	-0.281	1.057	1.137	0.256
Pan (2011)_02	SnR, Percent Body Fat, Curl-up(30s), Curl-up(60s), & PACER	0.670	0.240	0.058	0.200	1.140	2.791	0.005
		0.412	0.114	0.013	0.189	0.634	3.626	0.000

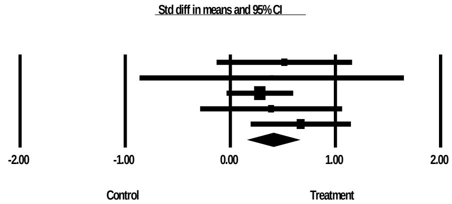


Figure 36. Forest Plot of Physiological & Biometric Indicator (adjusted PHY) – random effect model

*only the first author and publication year were shown

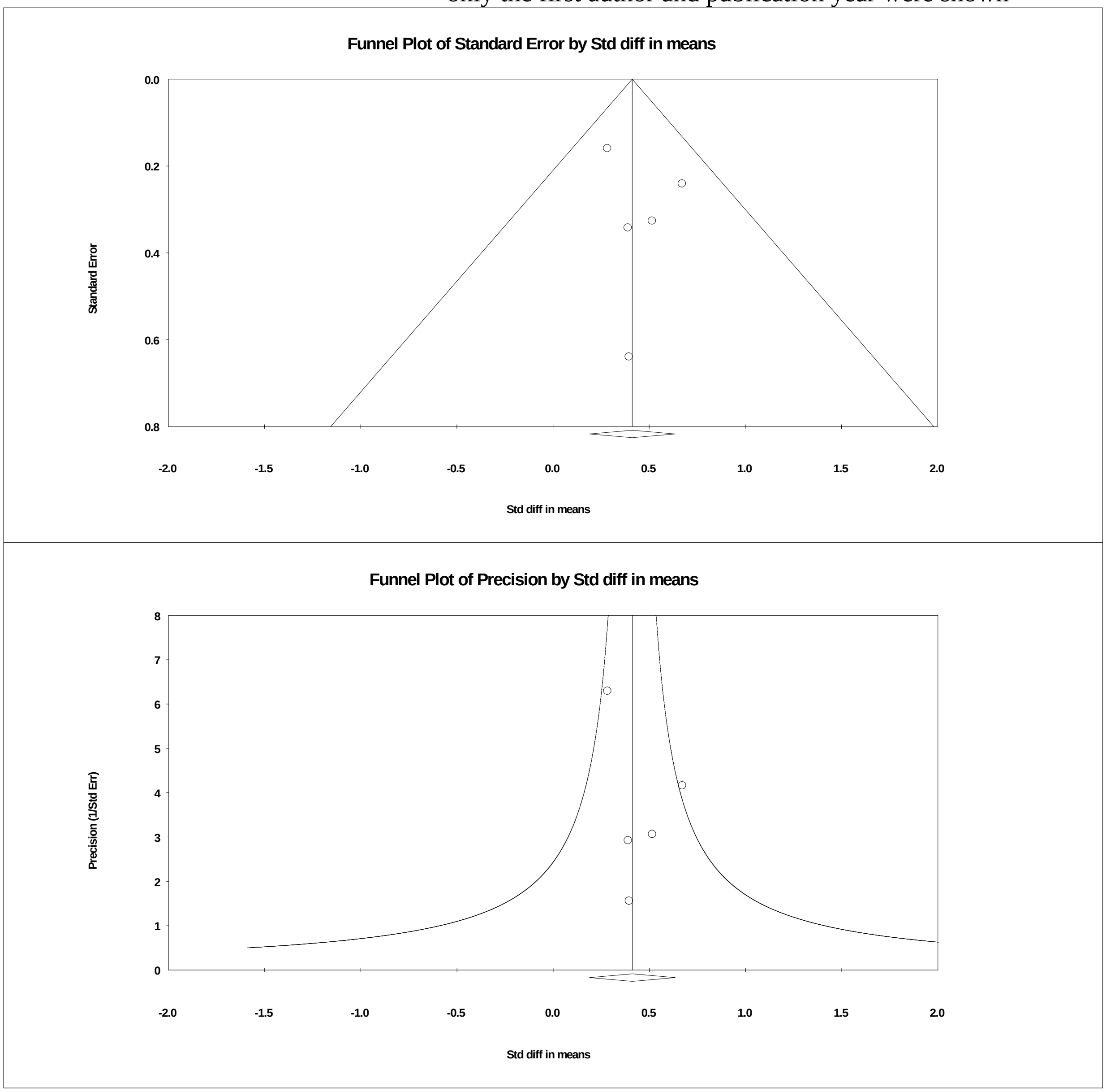


Figure 37. Funnel Plot of Physiological & Biometric Indicator (adjusted PHY) – random effect model

SOC adjusted

In the figure 38 and 39, a fixed effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size ($d = 0.296$, 95% CI = 0.108, 0.484, $p = 0.002$) showed the exercise gave a small to medium effect on enhancing performance in Social Cognition and Psychological Well-being (adjusted SOC) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

In the figure 40 and 41, a random effect model was applied, it showed a statistically significant result. The overall mean Cohen's effect size ($d = 0.505$, 95% CI = 0.152, 0.857, $p = 0.005$) showed the exercise gave a medium effect on enhancing performance in Social Cognition and Psychological Well-being (adjusted SOC) among children with ASD. The funnel plots showed a symmetrical shape, which indicated no publication bias was observed among the included studies.

For the outcome category of Social Cognition and Psychological Well-being (adjusted SOC), the heterogeneity Q statistic showed a statistical significance ($Q = 11.179$, df (Q) = 5, $p = 0.048$), hence, a random effect model was more appropriate in combining the data.

SOC adjusted (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Koening (2012)_01	ABC-Community-adaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_02	SSBS-2	0.775	0.367	0.135	0.056	1.495	2.113	0.035
Kern (2011)_02	TPCIS	0.088	0.121	0.015	-0.149	0.324	0.726	0.468
		0.296	0.096	0.009	0.108	0.484	3.084	0.002

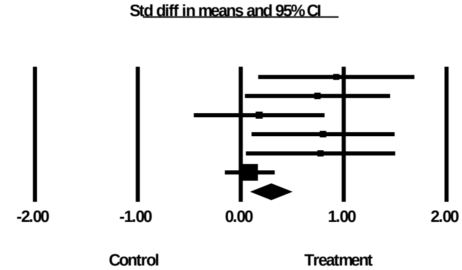


Figure 38. Forest Plot of Social Cognition & Psychological Well-being (adjusted SOC) – fixed effect model

*only the first author and publication year were shown

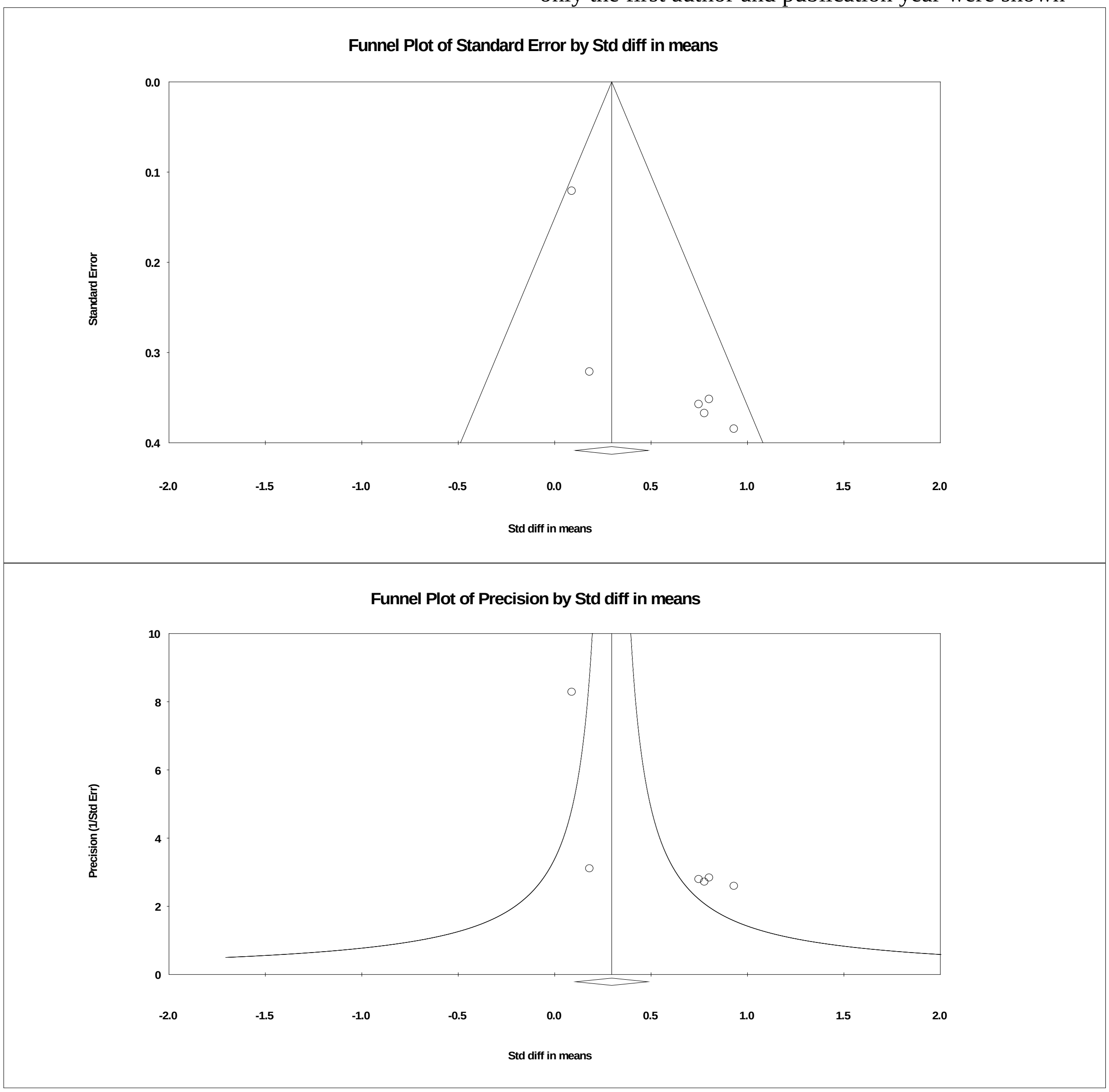


Figure 39. Funnel Plot of Social Cognition & Psychological Well-being (adjusted SOC) – fixed effect model

SOC adjusted (continue)

Study name	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Bahrami (2012)_01	SS-GARS-2	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Bass (2009)_02	SRS	0.746	0.357	0.127	0.046	1.446	2.089	0.037
Chan (2013)_03	SOC-ATEC	0.179	0.321	0.103	-0.450	0.809	0.559	0.576
Koening (2012)_01	ABC-Community-adaptive behavior	0.800	0.351	0.124	0.111	1.489	2.276	0.023
Pan (2010)_02	SSBS-2	0.775	0.367	0.135	0.056	1.495	2.113	0.035
Kern (2011)_02	TPCIS	0.088	0.121	0.015	-0.149	0.324	0.726	0.468
		0.505	0.180	0.032	0.152	0.857	2.805	0.005

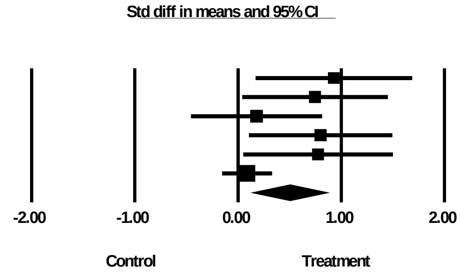


Figure 40. Forest Plot of Social Cognition & Psychological Well-being (adjusted SOC) – random effect model

*only the first author and publication year were shown

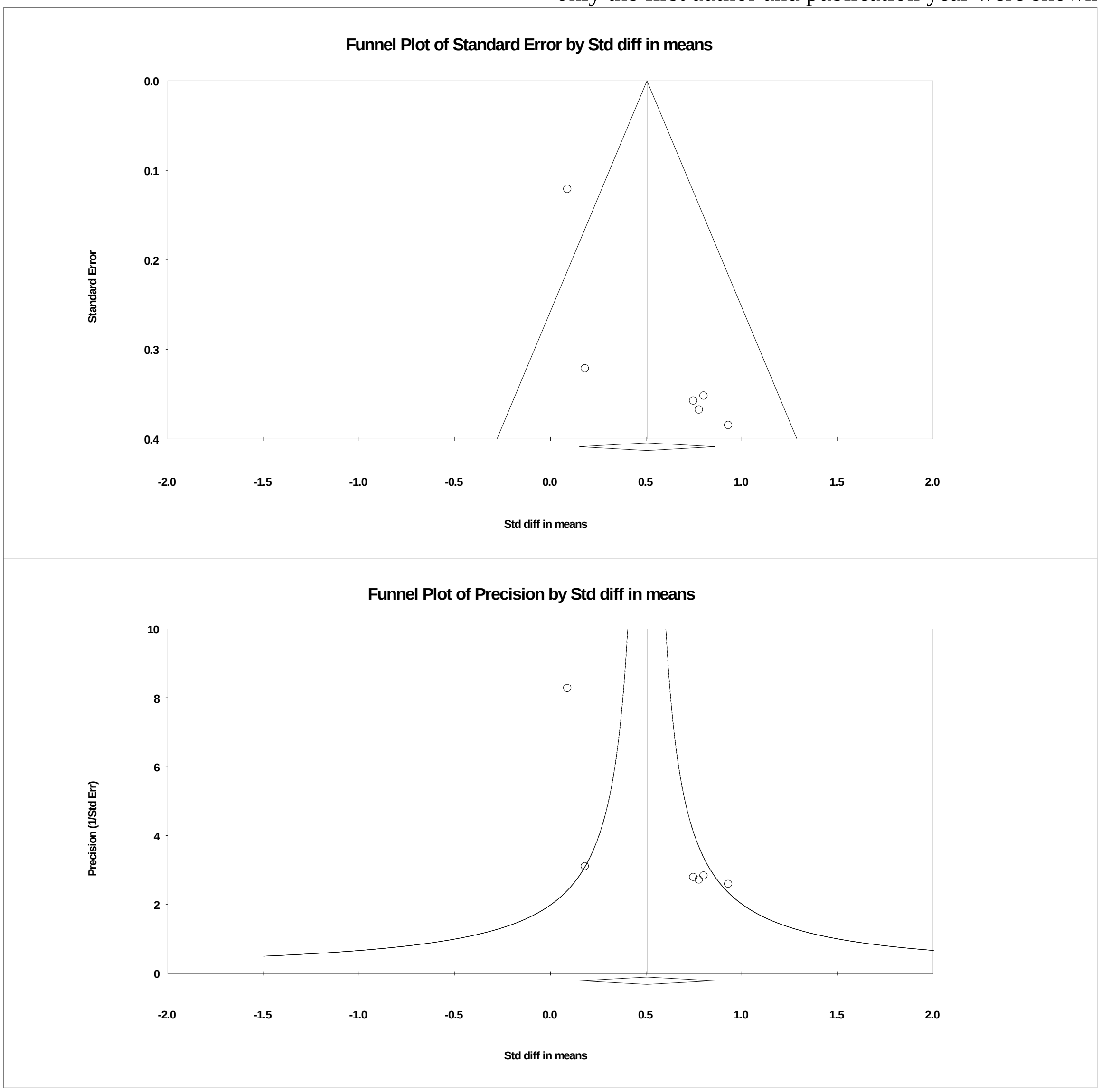


Figure 41. Funnel Plot of Social Cognition & Psychological Well-being (adjusted SOC) – random effect model

CHAPTER 4

Summary of Results and Discussion

Systematic review

The first step of the systematic review was the initial search. 6 university subscribed databases (4 in Hong Kong; 1 in United Kingdom; 1 in United Stated) and 6 online databases (Academica Sinica, PubMed (NLM), Web of Science (AHCI), Web of Science (Core Collection), Web of Science (SCI) and Web of Science (SSCI)) were included. The number of studies identified through online databases was 31,438 (41-12,286) which is much higher than those from universities, 885 (56-206).

After undergoing the screening procedure (delete duplications; identify relevance via titles and abstracts; read full-text articles), 11 studies were left. Those 11 selected studies were undergone the data extraction (extract and simplify the details of the studies, i.e., participants, age, place-environment, design, interventions/exercise components - type of exercise adopted, duration, intensity, frequency, outcome measures and findings) and critical appraisal (use the assessment tools - Jadad and modified CONSORT) procedures (details please see Figure 42).

The summary of findings for systematic review is contained in Table 11.

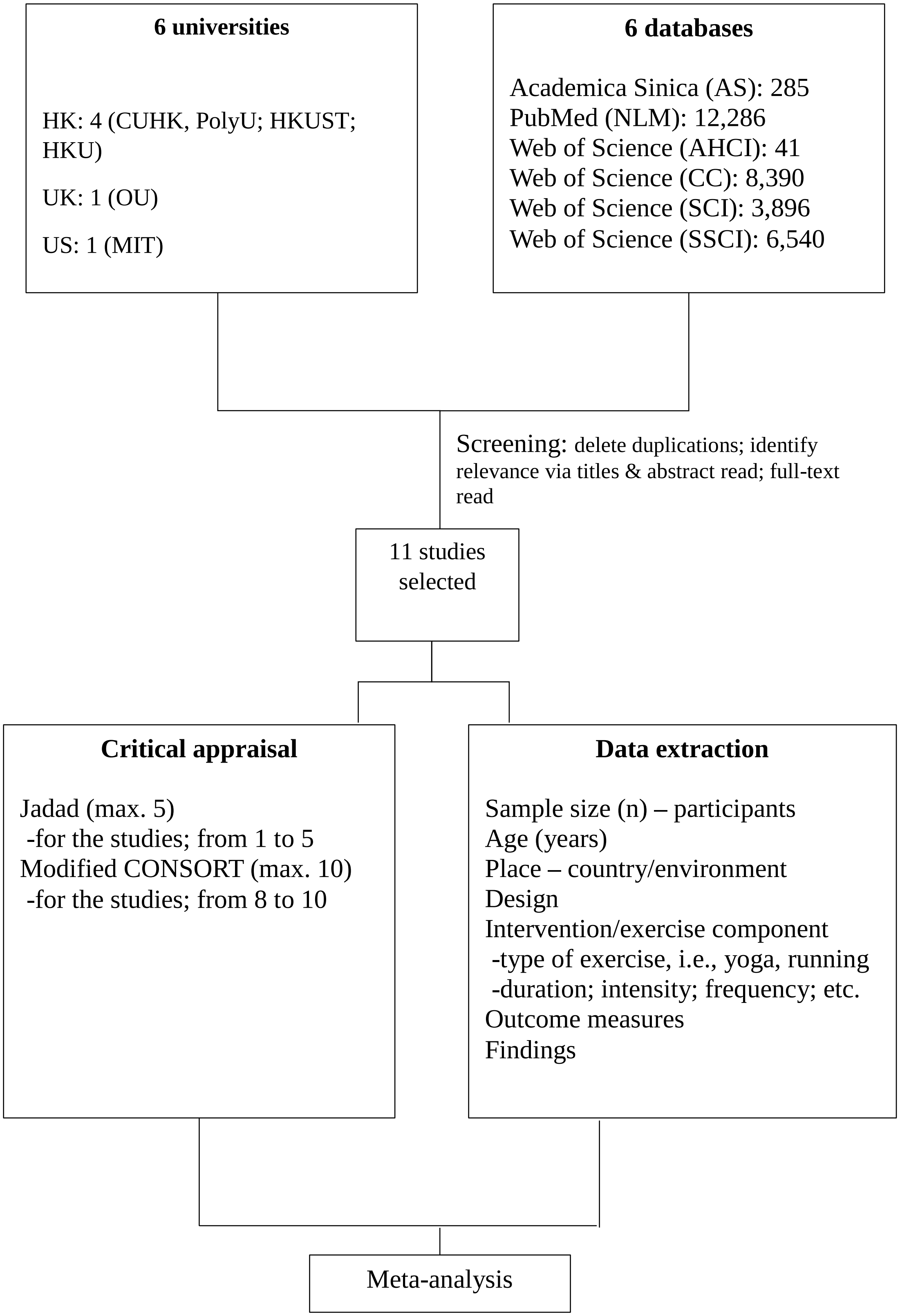


Figure 42. Flowchart of the systematic review

Table 11. Summary of findings

Study	**N** (M, F)	**Age** (M±SD)	**Place**	**Design**	**Intervention**	**Findings**
Bahrami et al. (2012)	30 (26, 4)	5-16 (T: 9.20±3.32; C: 9.06±3.33)	Iran	RCT	**Kata techniques training** (56sessions, 90 min/session; 4day/week,)	T > C** (SS-GARS-2)
Bass et al. (2009)	34 (29, 5)	4-10 (T: 6.95±1.67; C: 7.73±1.65)	US	RCT	**Horse riding** (1hr/session, 1session/week; 12 weeks)	T > C** (SP); T > C* (SRS)
Chan et al. (2013)	40 (36, 4)	6-17 (T: 11.28±3.90; C: 12.42±3.25)	HK	RCT	**Yoga** (1hr/session, 2session/week; 4weeks)	T > C** (PHY-ATEC; SOC-ATEC); T > C* (SEN-ATEC); N.S. (SPE-ATEC)
Fragala-Pinkham et al. (2008)	16 (11, 5)	6-12 (9.58±1.33)	US	A-B	**Aquatic exercise** (32-50min/session;2session/week; 14weeks)	T > C** (Half mile); N.S. (M-PEDI; FTS; Hip abductors; Knee extensors; Knee flexors; Ankle plantarflexors; Modified curl-up)
Fragala-Pinkham et al. (2011)	12 (11, 1)	6-12 (T: 9.60±2.60; C: 9.60±1.30)	US	CT	**Aquatic exercise** (40min/session; 2sessions/week; 14weeks)	T > C* (Swimming Classification Scale); N.S. (Mobility skills; 1/2 Mile; IPU; Modified curl-ups)
Kern et al. (2011)	24 (18, 6)	3-12 (7.80±2.9)	US	A-B	**Horse riding** (1hr/session; 1session/week; 6months)	T > C** (TPCIS -Mood and Tone); T > C* (SP-Auditory high-threshold; CARS); N.S. (TPCIS-Expressiveness; Responsiveness; Positive Regard; Negative Regard; Empathy; SP)
Koening et al. (2012)	46 (37, 9)	5-12 (T: 9.58; C: 8.58)	US	CT	**Yoga** (15-20 min/session; 5sessios/week; 16weeks)	T > C* (ABC-Community-maladaptive behavior)
Pan (2010)	16 (16, 0)	6-9 (T: 7.27±1.25; C: 7.20±.89)	Taiwan	CT	**Aquatic exercise** (20sessions; 90min/session; 2sessions/week)	T > C** (HAAR; AnB-SSBS-2); N.S. (SC-SSBS-2)
Pan (2011)	15 (15, 0)	7-12 (T: 9.31±1.67; C: 8.75±1.76)	Taiwan	CT	**Aquatic exercise** (28sessions; 60min/session; 2sessions/week)	T > C** (HAAR-IV); T > C* (HAAR-V; Physical fitness-Curl-ups 30s; Curl-ups 60s; N.S. (HAAR-III; Physical fitness-SnR; Percent body fat; 16-m PACER; BMI)
Pitetti et al. (2007)	10 (6, 4)	14-19 (T: 16.60±1.90; C: 17.40±1.10)	US	CT	**Treadmill walking** (individualized progression program; 9months)	N.S. (BMI)
Wuang et al. (2010)	60 (47, 13)	6-10 (M: 9.10±4.11; F: 9.11±3.76)	Taiwan	CT	**Horse riding** (1hr/session; 2sessions/week; 40sessions)	T > C** (BOTMP; TSIF)

Why those 11 included studies have small sample size?

It might cause by threefold. First, a small number of subjects can be quick to conduct with regard to enrolling participants, performing skill tests and reviewing the overall process. Second, the resources (i.e., administration cost, instructor, venue, etc.) required for a small number of subjects are relatively low, if the budget/funding was tight, it is better conducted in a small group of subjects rather than a large group. Third, it depends on the level of difficulties of recruiting the target group of subjects (autistic children), and the prevalence rate (of autism) around the community, other than that, the service or popularity of diagnosis and the privacy issue are also be the latent threads influencing the final sample size in those studies.

Why a wide variety of scales being used as outcome measures in those studies?

Since there is a wide range of outcomes under the interventions, and even wider range of outcome measures for researchers to select. On the other hand, the exercise content itself has its unique characteristics, for example, the kata technique's assessment will not be slightly different from the one for horse riding. It was reasonably suspected that in order to reach the program objectives and cater the individual needs probably, a wide variety of scales was hence being used in the currently included studies.

What kinds of exercise(s) being used more often? Why?

Horse riding and aquatic exercise were the two interventions being used more often. Three studies (Bass et al., 2009; Kern et al., 2011; Wuang et al., 2010) conducted by three different (groups of) researchers, whom adopted horse riding as an exercise intervention for autistic children; whereas four studies (Fragala-Pinkham et al., 2008; Fragala-Pinkham et al., 2011; Pan, 2010; Pan, 2011) conducted by two different (groups of) researchers, adopted the aquatic exercise as an intervention for children with autism. For horse riding, it requires the participant to learn how to communicate with animals, it will somehow train up some of their non-verbal communication skills, and help them understand the conditions, feelings, and thoughts of others. For aquatic exercise, it requires more mobility than the land activities, it also acts on stimulating the development of multi-sensory system (i.e., visual, audio, receptive perception, etc.). Those were believed are the possible reasons behind why choosing these two exercises (horse riding and aquatic exercise) more often.

Quality assessment of the selected studies - Jadad versus modified CONSORT

According to the results of the quality assessment of the included studies, most of the studies (54.54%; 6 out of 11) got only 1 mark according to the Jadad critical appraisal criteria; 4 of them (36.36%) got 3 marks; one study (Chan et al., 2013) got full mark. For the critical appraisal criteria of modified CONSORT, all of the 11 selected studies were over 7 (maximum 10) marks; 3 of them got 10 full marks; 6 of them got 9 marks; and the rest of 2 got 8 marks. Those discrepancies were due to the different criteria of the assessment tools. In Jadad scale, there are five basic requirement for assessing the quality of a specific control-trial study, each of the requirement weights for 20% (1 mark) of the total score (100%; full mark as 5), they are 1) randomization sounded (0/1); 2) randomization procedure described (0/1); 3) double-blinded (0/1); 4) double-blinded procedure described (0/1); and 5) withdrawals/dropouts described (0/1). Also, there are two deduction criteria: 1) inappropriate randomization process (0/-1); and 2) inappropriate double-blinded process (0/-1). Whereas, in modified CONSORT, there are ten requirement for assessing the quality of a clinical trial study, each of the requirement weights for 10% (1 mark) of the total score (100%; full mark as 10), they are 1) trial design (0/1); 2) participants (0/1); 3) sample size (0/1); 4) interventions (0/1); 5) outcomes (0/1); 60 randomization (0/1); 7) statistical methods (0/1); 8) primary objective (0/1); 9) program dosage (0/1); and 10) program variables (0/1). For this meta review study, since all of the included studies were not strictly followed the standardized randomized control trail design, as for the limitations of the educational context and the school/center/community based intervention settings, the use of Jadad scale might not be that appropriate for assessing the quality of those selected studies finally. It might somehow over-simplify the structure of what a "good" control-trial study should be, and resulted a relatively low score as using those standard criteria. On the other hand, the modified CONSORT provides more criteria based on the trial design, sampling issues, intervention details, i.e., program dosage, program variables, as well as statistical methods, and etc. Compared to the Jadad scale, the modified CONSORT might be a more comprehensive tool on assessing the study quality of that kind of intervention studies.

Additional part: theoretical framework of exercise interventions

Referring to each of the included studies [11 in the systematic review or 10 after excluding 1 outliner at the very end of meta-analysis (details will be discussed in the meta-analysis session)], none of them provided a theoretical framework of why exercise helps those individuals perform better on cognitive and/or motor tasks.

Children with autism might be diagnosed with intellectual disability (ID) [1/3 of autistic children with ID problem], information about theories adopted in studies of exercise interventions for children with ID might have relevance for exercise intervention studies targeting for children with autism. A recent systematic review conducted by Houwen, van der Putten, and Vlaskamp in 2014 reviewed the exercise type intervetions for people with ID, the results were coherent to this finding, they spotted out a similar trend, with a 45 included studies, very few studies (6/45) began with a sounded theoretical framework. The theories adopted were summarized in Table 11 below.

Table 12. Theoretical framework utilization

Theory utilized	**Number of studies**
Motor learning theory	2
Piaget's theory of sensory-motor intelligence development	1
Functional framework	1
Consolidation memory theory	1
Feuerstein's model of the 'active modifying environment' and Lebeer's 'stimulation ecology'	1
None	39
Total:	45

p.s. one study reported in more than one article, end up 45 in total

Motor learning theory was the most reported of theoretical basis, which was found in two studies (Choi, Meeuwsen, French, Sherrill, & McCabe, 2001; Choi, Meeuwsen, French, & Stenwall, 1999). Other studies reported using Piaget's theory of sensory-motor intelligence development (Pizzamiglio et al., 2008), consolidation memory theory

(McNeill & Mulholland, 2010), functional framework (Lancioni et al., 2003), and Feuerstein's model of the 'Active Modifying Environment' and Lebeer's 'Stimulation Ecology' (Lotan et al., 2012), correspondingly. Many studies implicitly mentioned theory without defining it as such. An assessment of the concepts behind the interventions used in studies that did not explicitly mention theory indicated that most of them relied on behavioral principles or behaviorism (Houwen, van der Putten, & Vlaskamp, 2014).

Why those included studies were not theory driven?

The data extraction form (Appendix B), designed according to the CRD (2009) guideline for undertaking reviews in Health Care, did not include the item of theory/theoretical framework of the study, and it was not a criterion in the quality assessment scales (Appendices C and D) either. It seems that the current studies adopted a practical approach, targeting at the individual needs of the participants rather than staying at the theoretical level of discussion, although those studies did not indicate what theory was adopted, many scholars (Houwen, van der Putten, & Vlaskamp, 2014; Pizzamiglio et al., 2008; Lotan et al., 2012) in the sociology field would categorize them as behaviorism (more practical, application oriented approach).

On the other hand, it was suspected that while preparing the final report, researchers might read the guideline for drafting their manuscript or instruction to authors, and those prestige journals had similar guidelines/instructions (as CRD did), normally, reporting the theory adopted was not a required item for assessing the original study/paper's quality since there is just a tiny little spacing for reporting the findings as well as its interpretations. All in all, the above-mentioned may also be the possible reason for not reporting the theory behind, bear in mind, it doesn't mean those included studies were not theory driven.

Suggested conceptual framework of a practical exercise program

Beyond the theoretical framework, the practical issue concerned is the program administration, as stated in the latest research of development of exercise program (Chan, Han, & Cheung, 2014; Houwen, van der Putten, & Vlaskamp, 2014), the common therapeutic exercise program includes 4 primary elements across 3 basic concerns. The 4 primary elements are participants, exercise components, environment, and intensity. The 3 basic concerns are theory, frequency, and outcome (Figure 43).

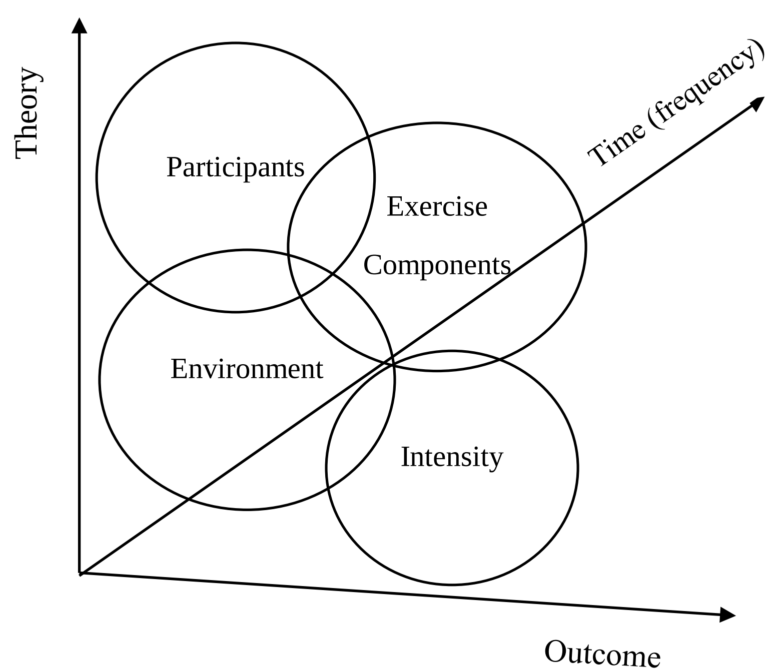


Figure 43. Conceptual framework of designing an exercise program

Strengths and weaknesses of the currently selected studies

The strength of the currently selected studies has threefold in general. The first one was including a clear statement of ASD definition at the very beginning. All the studies were clearly described the criteria of acceptance as their targeted subjects, the diagnostic confirmation (confirmed all the participants were ASD cases) would certainly enhance the validity of the studies. Second, the withdrawals or dropouts were described in a coherent manner in all 11 selected studies. It could be verified by the item 5 (withdrawals/dropouts described) scores of the Jadad scale among all the selected studies. Third, although a wide variety of measures was used among all the studies, those could specifically assess the target outcomes, and were tailor-made or wisely selected for the content of intervention/exercise itself, it could be easily spotted out from extracting the information of outcome measures for each of the studies.

The weakness of the identified studies has threefold also. Firstly, most of the included studies did not undergo the procedure of randomization. Only three of them (Bahrami et al., 2012; Bass et al., 2009; Chan et al., 2013) mentioned the randomization procedure under the study method sections, 8/11 of them were unknown (not mentioned), in that case, we could assume their samples were not randomized, since a randomization procedure could somehow buffer the selection bias, it was a criterion for evaluating study quality as well. Secondly, most of the included studies were not double-blinded, which means either the instructors or participants, or both of them, knew they are in control or experimental group, it would induce the experimenter and/or placebo effect(s) during the period. Finally, there is no standardized protocol of exercise, documented neither in the literature review nor the methodology part among all of the selected studies, each of the exercise intervention was instantly designed and was independent of different studies, even the same (group of) researcher(s), i.e., Fragala-Pinkham (2008) vs. Fragala-Pinkham (2011); Pan (2010) vs. Pan (2011), the exercise protocol in different years could be slightly different. It makes us difficult to compare the results of those intervention directly, since they were actually adopted different or slightly different protocol.

However, even different exercises were used in different studies, with not exactly the same outcome measures, the statistical method of meta-analysis, still, could provide a general picture, for instance, an influential implication for evaluation of those exercise-based interventions for children with autism. The following part would be continued on the discussion of quantifying and combining those intervention outcomes under the meta-analysis.

Meta-analysis

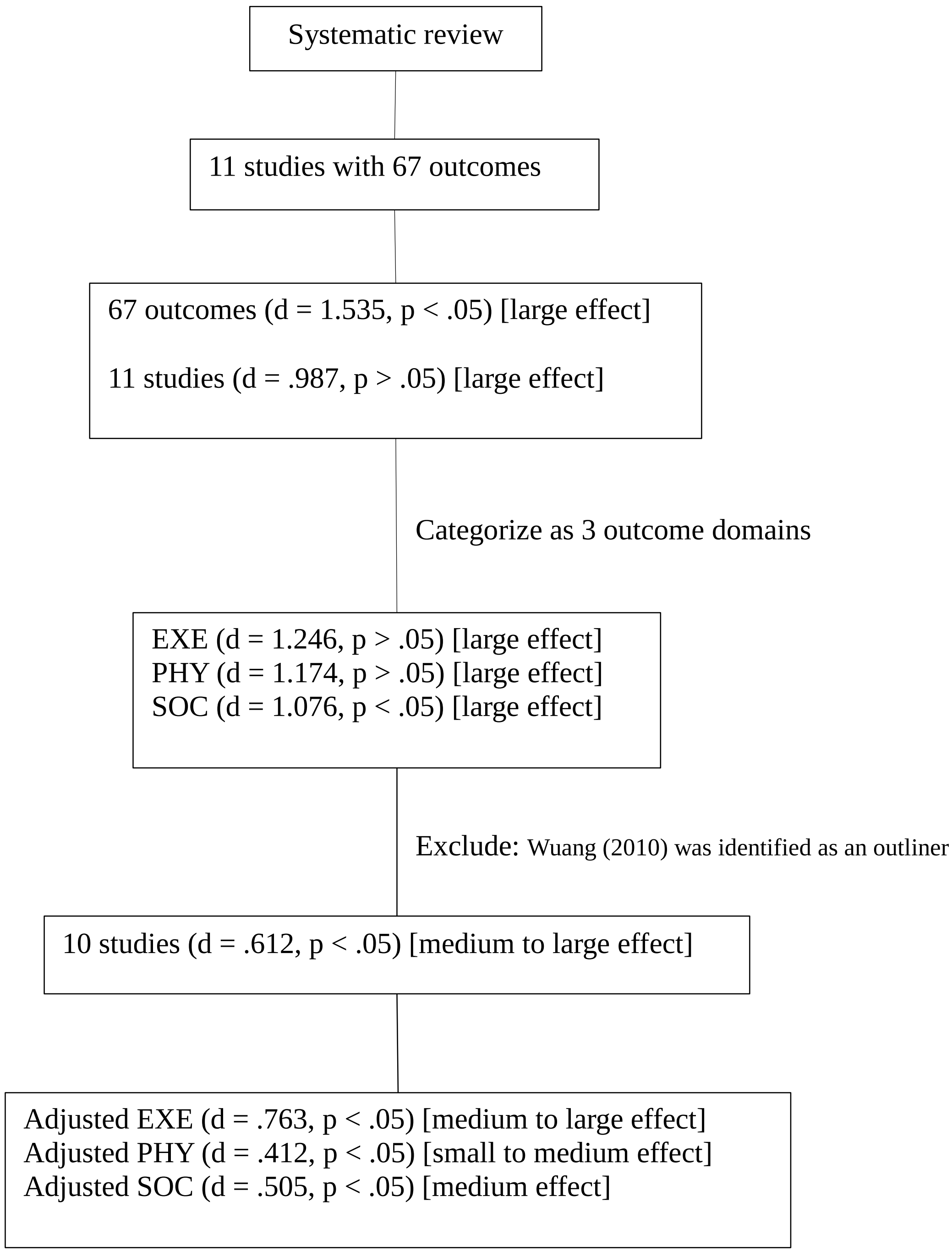
Eleven studies with 67 outcomes identified were further analyzed with statistical methods, called meta-analysis. The first analysis was performed for all 67 outcome measures extracted from those 11 selected studies, the overall effect size ($d = 1.535$, $p < .05$) showed a large effect with significance. While pooling those outcomes by studies, the overall effect size ($d = .998$, $p > .05$) showed still a large effect, but the result was not statistically significant. The inconsistent results stimulated us (the expert team) to explore for a comparatively appropriate outcome categorization method in order to better interpret the results, since grouping by studies might not be a suitable option due to the different characteristics or natures, such as rater discrepancies, between studies. After a further categorization with consensus of the expert team, involving four professionals – 1 occupational therapist, 1 educational psychologist, 2 speech therapists, and 1 physical educator (as the present researcher); three outcome categories were formed finally, which included Exercise Performance & Sport/Skill-related Fitness (EXE), Physiological & Biometric Indicator (PHY), and Social Cognition & Psychological Well-being (SOC).

The categorized outcome domains were undergone the analytical procedure for further investigation of the outcome effects. Throughout the exercise interventions, the outcome category of EXE showed a large effect but statistically insignificant ($d = 1.246$, $p > .05$), the PHY showed a large effect but also statistically insignificant ($d = 1.174$, $p > .05$), the only one showed a large effect with statistically significant was SOC ($d = 1.076$, $p < .05$).

Two out of three outcome domains showed a statistically insignificant results. Why did it happen? After reviewing all those categorized forest plots and the 11 studies' results, one study (Wuang et al., 2010) was identified as an outliner among all selected studies. In order to remove the effect from outliner(s), this study (Wuang et al., 2010) was finally decided to exclude for the later part of the analysis.

After excluding the outliner study (Wuang et al., 2010), the remaining 10 studies were underwent the analytical procedure again. By comparing the effects between 11 studies and 10 studies, an effect size calculation of the pooled outcomes by 10 studies was performed, and it showed a medium to large effect with significance ($d = .612$, $p < .05$), whereas the pooled outcomes by 11 studies showed a large effect without significance ($d = .998$, $p > .05$).

While excluding the outliner study (Wuang et al., 2010), the adjusted outcome categories were also underwent the analytical procedure for investigating the exercise effects for children with ASD. First, the outcome category of EXE showed a medium to large effect with significance ($d = .763$, $p < .05$). Second, the PHY showed a small to medium effect with significance ($d = .412$, $p < .05$). Third, the SOC showed a medium effect, also with statistical significance ($d = .505$, $p < .05$). All in all, those adjusted outcome domains generally showed an average medium effect on the exercise interventions, and those effects were all statistically significant (see Figure 44).



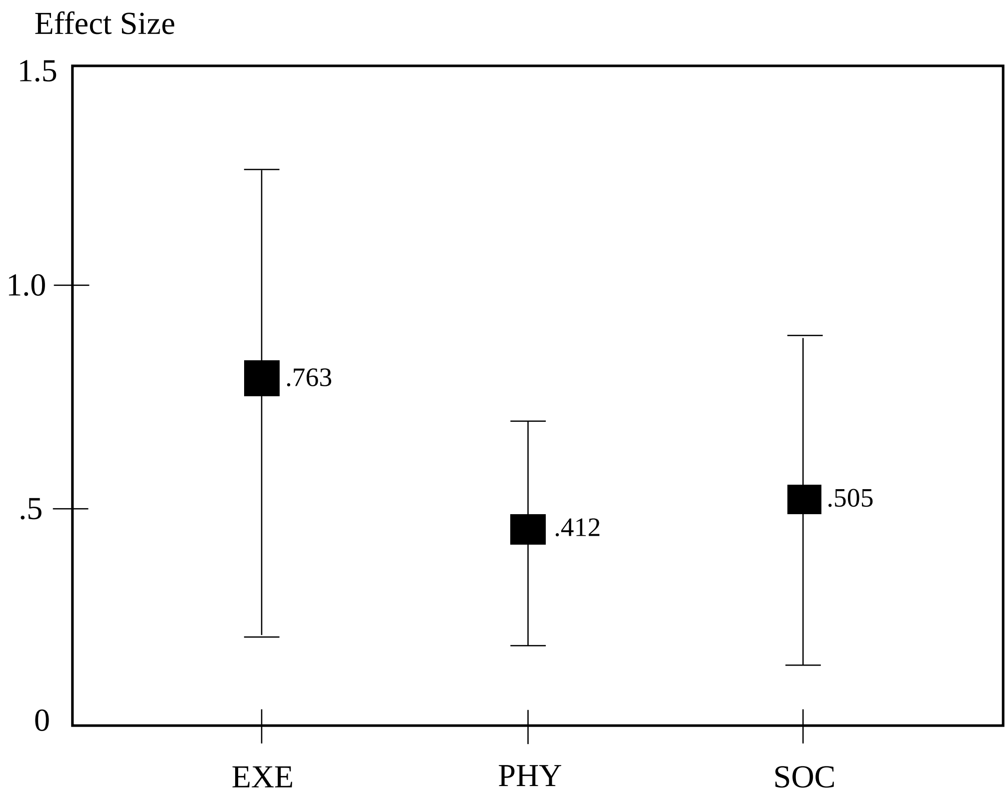
Note: Exercise performance & sport/skill-related fitness (EXE)
Physiological & biometric indicator (PHY)
Social cognition & psychological well-being (SOC)

Figure 44. Flowchart of the meta-analysis

Effects of the exercise interventions

Evidence was found effects on exercise programs for children with ASD. In this study, ten studies (excluded one which identified as outliner) were finally included for analysis and the effect sizes of each study were computed. The overall effect was medium to large, and since the measures among those studies were different in nature, in order to avoid "adding apples and oranges" problem, three groups of outcomes were identified by the consensus of a team of experts, abovementioned, in the field of sub-medical profession, physical education and sports science, and educational psychology, these three outcome measure groups are 1) Exercise Performance & Sport/Skill-related Fitness (EXE); 2) Physiological & Biometric Indicator (PHY); and 3) Social Cognition & Psychological Well-being (SOC), hence, the results in this study were interpreted separately according to the specific outcome domain it belongs to. As for the abovementioned, those three outcome domains showed an average medium effect (from .412 to .763) with significance ($p < .05$) in statistics (see Figure 45).

The results were aligned to the previous review studies (Lang et. al., 2010; Petrus et. al., 2008; Sorensen & Zarrett, 2008; Sowa & Meulenbroek, 2011), moreover, the recent study quantified those outcomes by measures, by studies, as well as by theoretical measure domains, which, to a certain extent, gave a more objective conclusion and provided a relatively holistic picture for the purpose of program evaluation. Those are the advantages of systematic review followed by a meta-analysis type studies.



Note: Exercise performance & sport/skill-related fitness (EXE)
Physiological & biometric indicator (PHY)
Social cognition & psychological well-being (SOC)

Figure 45. Effects of exercise interventions by three outcome domains

For some highlighted issues, such as publication bias, model selection for the meta-analysis, and etc., are discussed below.

Issue of publication bias

It is undeniable that positive results may favor for publications. Publication bias is a crucial issue since it relates to the validity of an enormous pool of published studies. To assess the publication bias, indicators, such as Kendall's tau (Begg & Mazumdar, 1994), Harbold-Egger's N-false rate (Harbord, Egger, & Sterne, 2006), and etc. The most common one is establishing a funnel plot (Egger, Smith, Schneider, & Minder, 1997) for detecting the distribution trend among the selected studies (effect size against standard error). In this study, the measures were pooled by studies, and finally categorized by outcome domains, applying funnel plots for assessing the publication bias, by observing the shape (symmetric or asymmetric; for asymmetric, right-skewed or left-skewed) of those plots, was operationally feasible. The polished outcomes by 10 selected studies and the final categorized outcomes by three domains (EXE, PHY, and SOC) were all showed a symmetrical shape, which implied no observable publication bias among studies.

Model selection

For selecting the statistical model of computing the overall effect size among measures/studies, the heterogeneity indicator, Cochran Q was computed. Cochran Q follows a chi-squared distribution, and a statistically significant Q value indicates that there is an extensive amount of variability across the included studies, a random effect model should be adopted rather than the fixed effect model. It is a key criterion for model selection in this study. The existing results showed that all of the outcome categories exhibited a significant Q, excepting the adjusted EXE domain which could adopted the remaining fixed effect model, which indicated a random effect model was more appropriate for calculating the effect sizes of those outcomes since the heterogeneity among studies was identified.

Table 13. Selection of fixed/random models

	Fixed effect model	**Random effect model**
67 measures		✓**
11 studies		✓
EXE		✓
PHY		✓
SOC		✓*
10 studies		✓**
EXE (adjusted)		✓**
PHY (adjusted)	✓**	
SOC (adjusted)		✓**

*p < .05; **p < .01

Note: Exercise performance & sport/skill-related fitness (EXE)
Physiological & biometric indicator (PHY)
Social cognition & psychological well-being (SOC)

Even with a random effect model, the outcomes by 11 selected studies, by unadjusted EXE, and by unadjusted PHY, were showed a statistically insignificant effect, which means no effects on exercise interventions throughout those categories. However, after excluding one outliner study, the remaining 10 selected studies and the 3 adjusted outcome domains showed a statistically significant positive effects on exercise interventions for children with ASD (see Table 12).

Sampling size estimation for the follow-up/future study

As for the continuing investigation of the exercise effect on children with ASD, the currently obtained effect sizes could be used to calculate the require sample size for the similar intervention study as the follow-up/future research do. For example, in this study, the overall effect sizes were .763, .412, and .505, for EXE, PHY, and SOC respectively;

set the alpha level as .05, power as .80, number of groups as 2 (exercise vs control), the minimum sample size required ranges from 16 to 38. Which means if you would like to conduct a similar control-trail designed study, a minimum sample of 38 (pick the highest one among the three outcome domains) was considered on drafting the research plan ahead (see Table 13).

Table 14. Sample size estimation

Domain	Effect size	α	Power (1-β)	Number of groups	Required **sample size** (n)
EXE	.763	.05	.80	2	16
PHY	.412	.05	.80	2	38
SOC	.505	.05	.80	2	30

EXE: Exercise performance & sport/skill-related fitness
PHY: Physiological & biometric indicator
SOC: Social cognition & psychological well-being

What type(s) of exercises exhibited the highest score(s) of effect? (Which study contributed most for the overall effect of exercise interventions?)

According to the results in the meta-analysis section, the study by Wuang et al. (2010) exhibited the highest score (4.529) of effect (Figure 10-13), however, this study was excluded from the later part of the analysis since its oversensitive measures (BOTMP and TSIF testing batteries) comparing to the others, if keep it for the analysis, the overall effect size would be dominated by those measures, this study would weight a highest proportion in the final score of lumped effect, hence it should be seemed as an outline in that case (see Figure 46).

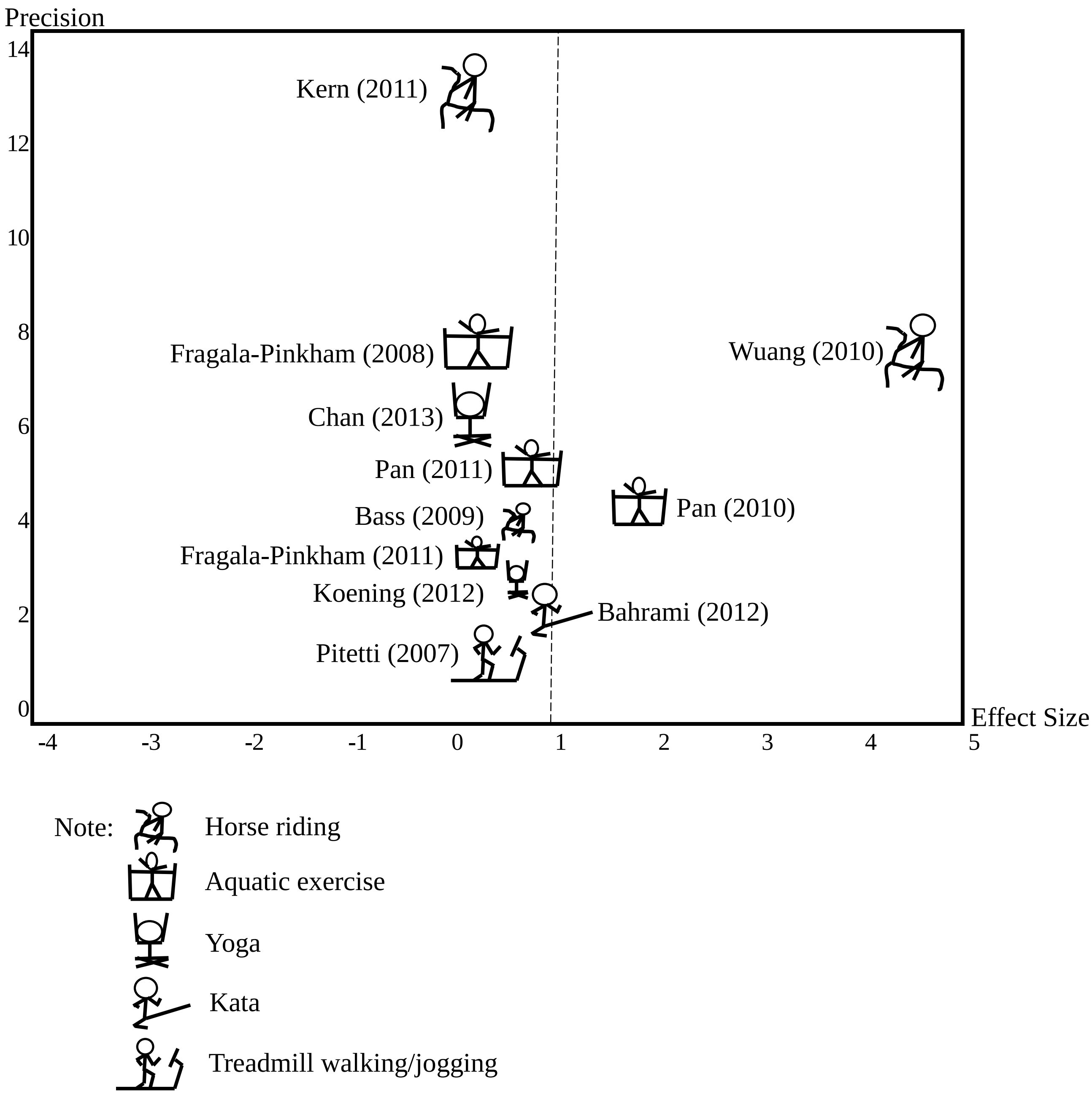


Figure 46. Precision versus Effect Size

If the study by Wuang et al. (2010) was excluded, the aquatic exercise study by Pan (2010) was the one with highest score in effect size. In order to answer the question “what type of exercise exhibited the highest effect?” based on the current results, we could conclude that the aquatic exercise might be a type of exercises which exhibited a highest score (1.593) of effect comparing to other types of exercise (i.e., Kata, Yoga, etc.). If the study by Wuang et al. (2010) was included, since the type of exercise which Wuang et al. (2010) adopted for was horsing riding, the highest score (4.529) of effect type exercise should be horsing riding. However, the lowest score (.150) one (Kern et al., 2011) was also selected horsing riding as its exercise intervention. We could not make a judge by the current results for considering what type of exercise was the most effective one while planning an exercise intervention ahead or in the future studies. A potential threat was the variation of the outcome measures (different studies adopted different measurement tools), it would seriously affect the final effect unless all the studies included were used the exactly same measurement tool or kit and the testing protocol was highly identical, then the effect of this threat could be minimized. However, for the current selected studies, even it was conducted by the same researchers, i.e., Pan (2010) versus Pan (2011), Fragala-Pinkham (2008) versus Fragala-Pinkham (2011), and adopted the similar exercise interventions (aquatic exercise), the outcome measures utilized were different! It might be due to the different focuses or considerations while making such changes by those researchers. Thus, reaching the objective to spot out the most effective exercise for helping children with autism might not be that feasible.

Recommendations on exercise characteristics (for children with autism) synthesized from the recent studies?

After synthesizing the extracted information from the selected studies, the characteristics of exercise intervention (i.e., duration, frequency, and session length) were generally 12-14 weeks, 1-2 sessions per week, and 60-90 minutes per session. However, if we adopted those synthesized characteristics as guidelines or recommendations for planning the future research or exercise program ahead, the interpretations must be cautious. As for a Yoga lesson, 60-90 minutes in length might be understandable, since the exercise intensity might not be that high enough and the children were plausibly manageable, if for a treadmill jogging/running or swimming lesson, 60-90 minutes might be too harsh for them to work out. On the other hand, all the researchers in those 11 selected studies did not specify the time allocation of each session, for example, in an exercise intervention of horse riding, the session length given of three included studies (Bass et al., 2009; Kern et al., 2011; Wuang et al., 2010) were 60 minutes, however, we did not know how much time the child was actually riding on the horse, since the preparation, debriefing and instruction time might be varied, and the exact time of horse riding might be less than 15-30 minutes, similar circumstances might happen in other exercise interventions as well. Therefore, it was conceivable that the nature of the exercise might be one of the crucial factors affecting the duration, frequency and session length of the intervention, still, the characteristics synthesized from the recent studies could be seemed as a reference rather than a conclusive guideline (i.e., golden rule).

Limitations

In this study, four types of studies were excluded, which are 1) non-control-trial studies; 2) non-English studies; 3) unpublished articles (conference papers); and 4) studies before 2000. Those excluded studies might have certain weightings and might change the results of the later part of this study - meta-analysis.

On the other hand, the program duration, frequency, and the length of each session, abovementioned, would be the potential factors affecting the overall effectiveness. However, since relatively few studies included in this study comparing to Shin and Park (2012) (11 in this study versus 14 in Shin & Park, 2012), although most of them had similar duration, frequency, and session length, i.e., 12-14 weeks, 2 session/week, and 60-90 minutes, the characteristics between ASD and ID kids are not entirely the same. Moreover, within 60-90 minutes exercise session, the actual exercising time of the participants did was untold, and the time allocations of each session, i.e., engagement time, management time, off-task time, and etc. were unknown as well. One study (Pitetti, 2007) did not report even the frequency and session length since it was an individualized progression program. For those reasons, the split-up sub-outcome analysis hence might not be appropriate, since it would over empathize the impact with a relatively few information as well as the wide varied exercise nature, i.e., the experience of horse riding could be very different from Kata training.

Furthermore, the nature and the mode of exercise, and the age range of the participants would also be the factors influencing the outcome; as exercises varied across studies, and the current focus was on children with ASD, but the age range did vary among, from 5 to 19 in these studies, which probably covered some of the adolescences or even adults instead of children. However, those were the solely available articles from institutional and online databases with the recent search keywords and screening method, a review of the current searching and screening procedure was recommended to be conducted in future.

Conclusion

For the systematic review, 11 control-trial studies, adopting exercise as intervention for children with autism, were identified from the recently accessible institutional and international e-databases; the exercises involved in the studies including horse riding, aquatic exercise, Yoga, Kata, and treadmill walking/jogging; each exercise intervention generally lasted for 12-14 weeks, 1-2 sessions per week, and 60-90 minutes per session; under the critical appraisal, those identified studies exhibited a fair (average) quality.

For the meta-analysis, current research evidence revealed a positive effect on exercise interventions for children with autism among three outcome categories: 1) Exercise Performance & Sport/Skill-related Fitness (EXE); 2) Physiological & Biometric Indicator (PHY); and 3) Social Cognition & Psychological Well-being (SOC). Further studies were desirable to confirm the moderating mechanism behind; might be deemed as crucial.

REFERENCES

Anderson, A., Moore, D. W., Godfrey, R., & Fletcher-Flinn, C. M. (2004). Social skills assessment of children with autism in free-play situations. *Autism, 8*(4), 369-385. doi: 10.1177/1362361304045216

Anderson, J. A., & Hinton, G. E. (1981). Models of information processing in the brain. *Parallel models of associative memory*, 9-48.

American Psychological Association (2013). *Diagnostic and statistical manual of mental disorders, (DSM-5®)*: American Psychiatric Pub.

Bahrami, F., Movahedi, A., Marandi, S. M., & Abedi, A. (2012). Kata techniques training consistently decreases stereotypy in children with autism spectrum disorder. *Research in developmental disabilities, 33*(4), 1183-1193.

Baker, D. L. (2007). Defining autism in Canada: Unfolding the public aspects of neurological disability. *The Social Science Journal, 44*(4), 687-697. doi: http://dx.doi.org/10.1016/j.soscij.2007.10.010

Baranek, G. (2002). Efficacy of Sensory and Motor Interventions for Children with Autism. *Journal of Autism and Developmental Disorders, 32*(5), 397-422. doi: 10.1023/A:1020541906063

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition, 21*(1), 37-46. doi: http://dx.doi.org/10.1016/0010-0277(85)90022-8

Bass, M., Duchowny, C., & Llabre, M. (2009). The Effect of Therapeutic Horseback Riding on Social Functioning in Children with Autism. *Journal of Autism and Developmental Disorders, 39*(9), 1261-1267. doi: 10.1007/s10803-009-0734-3

Bauman, M. L., & Kemper, T. L. (2005). *Structural brain anatomy in autism: what is the evidence*: The Neurobiology of Autism. 2nd ed. Baltimore, MD: Johns Hopkins University Press.

Bauminger, N. (2002). The Facilitation of Social-Emotional Understanding and Social Interaction in High-Functioning Children with Autism: Intervention Outcomes. *Journal of Autism and Developmental Disorders, 32*(4), 283-298. doi: 10.1023/A:1016378718278

Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics, 50*(4), 1088-1101. doi: 10.2307/2533446

Belfer, M. L. (2008). Child and adolescent mental disorders: the magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry, 49*(3), 226-236. doi: 10.1111/j.1469-7610.2007.01855.x

Berkeley, S. L., Zittel, L. L., Pitney, L. V., & Nichols, S. E. (2001). Locomotor and object control skills of children diagnosed with autism. *Adapted Physical Activity Quarterly, 18*(4), 405-416.

Bhat, A. N., Landa, R. J., & Galloway, J. C. (2011). Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders. *Physical Therapy, 91*(7), 1116-1129. doi: 10.2522/ptj.20100294

Blakemore-Brown, L. (2002). *Reweaving the autistic tapestry: Autism, Asperger's syndrome, and ADHD*: Jessica Kingsley.

Bouras, N., Holt, G., Day, K., & Dosen, A. (1999). *Mental health in mental retardation: The ABC for mental health, primary care and other professionals.* Paper presented at the World Psychiatric Association.

Braff, D. L. (1993). Information Processing and Attention Dysfunctions in Schizophrenia. *Schizophrenia Bulletin, 19*(2), 233-259. doi: 10.1093/schbul/19.2.233

Chakrabarti, S., & Fombonne, E. (2001). PErvasive developmental disorders in preschool children. *JAMA, 285*(24), 3093-3099. doi: 10.1001/jama.285.24.3093

Chakrabarti, S., & Fombonne, E. (2005). Pervasive Developmental Disorders in Preschool Children: Confirmation of High Prevalence. *American Journal of Psychiatry, 162*(6),

1133-1141. doi: doi:10.1176/appi.ajp.162.6.1133
Chan, A., Han, Y., & Cheung, M.-c. (2014). Chinese Chan-Based Prospective Neuropsychological Intervention for Autistic Children. In V. B. Patel, V. R. Preedy, & C. R. Martin (Eds.), *Comprehensive Guide to Autism* (pp. 2333-2355): Springer New York.
Chan, A. S., Sze, S. L., Siu, N. Y., Lau, E. M., & Cheung, M. C. (2013). A chinese mind-body exercise improves self-control of children with autism: a randomized controlled trial. *PLoS One, 8*(7), e68184. doi: 10.1371/journal.pone.0068184
Choi, S. O., Meeuwsen, H. J., French, R., Sherrill, C., & McCabe, R. (2001). Motor skill acquisition, retention, and transfer in adults with profound mental retardation. *Adapted Physical Activity Quarterly*, *18*(3), 257-272.
Choi, S., Meeuwsen, H. J., French, R., & Stenwall, J. (1999). Learning and control of simple aiming movements by adults with profound mental retardation. *Adapted Physical Activity Quarterly, 16*, 167-177.
Cochran, W. G. (1954). Some Methods for Strengthening the Common χ2 Tests. *Biometrics, 10*(4), 417-451. doi: 10.2307/3001616
Cohen, J. (1988). *Statistical power analysis: A computer program*: Routledge.
Control, C. f. D. (2009). Prevention.(2009). *The guide to community preventive services glossary. Retrieved December, 6*.
Cornish, K. M., & McManus, I. C. (1996). Hand preference and hand skill in children with autism. *Journal of Autism and Developmental Disorders, 26*(6), 597-609. doi: 10.1007/BF02172349
Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin, 104*(2), 163.
Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology, 40*(2), 271.
de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship Between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) Classification in Children and Adolescents with Mental Retardation. *Journal of Autism and Developmental Disorders, 34*(2), 129-137. doi: 10.1023/B:JADD.0000022604.22374.5f
Downey, R., & Rapport, M. J. K. (2012). Motor Activity in Children With Autism: A Review of Current Literature. *Pediatric Physical Therapy, 24*(1), 2-20. doi:10.1097/PEP.0b013e31823db95f
Duffy, E. (1962). *Activation and behavior*. Oxford, England: Wiley.
Durkin, M. (2002). The epidemiology of developmental disabilities in low-income countries. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*(3), 206-211. doi: 10.1002/mrdd.10039
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). *Bias in meta-analysis detected by a simple, graphical test* (Vol. 315).
Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., . . . Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Research, 5*(3), 160-179. doi: 10.1002/aur.239
Emck, C., Bosscher, R., Beek, P., & Doreleijers, T. (2009). Gross motor performance and self-perceived motor competence in children with emotional, behavioural, and pervasive developmental disorders: a review. *Developmental medicine & child neurology, 51*(7), 501-517. doi:10.1111/j.1469-8749.2009.03337.x
Fatemi, S. H., Merz, A., & Realmuto, G. (2003). The Roles of Reelin, Bcl2, and Serotonin in Cerebellar Pathology in Autism. *Journal of Developmental and Physical Disabilities,*

15(1), 1-22. doi: 10.1023/A:1021457519382
Fournier, K., Hass, C., Naik, S., Lodha, N., & Cauraugh, J. (2010). Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis. *Journal of Autism and Developmental Disorders, 40*(10), 1227-1240. doi: 10.1007/s10803-010-0981-3
Fragala-Pinkham, M., Haley, S. M., & O'Neil, M. E. (2008). Group aquatic aerobic exercise for children with disabilities. *Developmental Medicine and Child Neurology, 50*(11), 822-827.
Fragala-Pinkham, M. A., Haley, S. M., & O’Neil, M. E. (2011). Group swimming and aquatic exercise programme for children with autism spectrum disorders: A pilot study. *Developmental Neurorehabilitation, 14*(4), 230-241. doi: doi:10.3109/17518423.2011.575438
Ganz, M. L. (2007). THe lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics & Adolescent Medicine, 161*(4), 343-349. doi: 10.1001/archpedi.161.4.343
Gentile, I., Zappulo, E., Militerni, R., Pascotto, A., Borgia, G., & Bravaccio, C. (2013). Etiopathogenesis of autism spectrum disorders: Fitting the pieces of the puzzle together. *Medical Hypotheses, 81*(1), 26-35. doi: http://dx.doi.org/10.1016/j.mehy.2013.04.002
Gidley Larson, J. C., Bastian, A. J., Donchin, O., Shadmehr, R., & Mostofsky, S. H. (2008). *Acquisition of internal models of motor tasks in children with autism* (Vol. 131).
Gillberg, C. (2002). *A guide to Asperger syndrome*: Cambridge University Press.
Glazebrook, C., Gonzalez, D., Hansen, S., & Elliott, D. (2009). The role of vision for online control of manual aiming movements in persons with autism spectrum disorders. *Autism, 13*(4), 411-433. doi: 10.1177/1362361309105659
Harbord, R. M., Egger, M., & Sterne, J. A. C. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine, 25*(20), 3443-3457. doi: 10.1002/sim.2380
Hedges, L. V., & Olkin, I. (2014). *Statistical method for meta-analysis*: Academic press.
Hewetson, A. (2002). *The stolen child: Aspects of autism and Asperger syndrome*: Bergin & Garvey Westport, CT.
Hill, A., & Spittlehouse, C. (2001). *What is critical appraisal?* : Hayward Medical Communications.
Hoeksma, M., Kemner, C., Verbaten, M., & van Engeland, H. (2004). Processing Capacity in Children and Adolescents with Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders, 34*(3), 341-354. doi: 10.1023/B:JADD.0000029555.98493.36
Houwen, S., van der Putten, A., & Vlaskamp, C. (2014). A systematic review of the effects of motor interventions to improve motor, cognitive, and/or social functioning in people with severe or profound intellectual disabilities. *Research in Developmental Disabilities, 35*(9), 2093-2116. doi: http://dx.doi.org/10.1016/j.ridd.2014.05.006
Jadad, A. R., Cook, D. J., Jones, A., & et al. (1998). Methodology and reports of systematic reviews and meta-analyses: A comparison of cochrane reviews with articles published in paper-based journals. *JAMA, 280*(3), 278-280. doi: 10.1001/jama.280.3.278
Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials, 17*(1), 1-12. doi: http://dx.doi.org/10.1016/0197-2456(95)00134-4
Jankowicz-Szymanska, A., Mikolajczyk, E., & Wojtanowski, W. (2012). The effect of physical training on static balance in young people with intellectual disability. *Research in Developmental Disabilities, 33*(2), 675-681. doi: http://dx.doi.org/10.1016/j.ridd.2011.11.015
Jansiewicz, E., Goldberg, M., Newschaffer, C., Denckla, M., Landa, R., & Mostofsky, S. (2006).

Motor Signs Distinguish Children with High Functioning Autism and Asperger's Syndrome from Controls. *Journal of Autism and Developmental Disorders, 36*(5), 613-621. doi: 10.1007/s10803-006-0109-y

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251. doi: 10.1037/h0034747

Kamijo, K., Nishihira, Y., Hatta, A., Kaneda, T., Wasaka, T., Kida, T., & Kuroiwa, K. (2004). Differential influences of exercise intensity on information processing in the central nervous system. *European Journal of Applied Physiology, 92*(3), 305-311. doi: 10.1007/s00421-004-1097-2

Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological bulletin, 70*(6, Pt.1), 387-403. doi: 10.1037/h0026739

Kern, J. K., Fletcher, C. L., Garver, C. R., Mehta, J. A., Grannemann, B. D., Knox, K. R., . . . Trivedi, M. H. (2011). Prospective trial of equine-assisted activities in autism spectrum disorder. *Alternative therapies in health and medicine*(17), 14-20.

King, M., & Bearman, P. (2009). Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology, 38*(5), 1224-1234. doi: 10.1093/ije/dyp261

King, M. D., & Bearman, P. S. (2011). Socioeconomic Status and the Increased Prevalence of Autism in California. *American Sociological Review, 76*(2), 320-346. doi: 10.1177/0003122411399389

King, M. D., Fountain, C., Dakhlallah, D., & Bearman, P. S. (2009). Estimated Autism Risk and Older Reproductive Age. *American journal of public health, 99*(9), 1673-1679. doi: 10.2105/AJPH.2008.149021

Klin, A., Volkmar, F. R., Sparrow, S. S., Cicchetti, D. V., & Rourke, B. P. (1995). Validity and Neuropsychological Characterization of Asperger Syndrome: Convergence with Nonverbal Learning Disabilities Syndrome. *Journal of Child Psychology and Psychiatry, 36*(7), 1127-1140. doi: 10.1111/j.1469-7610.1995.tb01361.x

Koenig, K. P., Buckley-Reen, A., & Garg, S. (2012). Efficacy of the Get Ready to Learn Yoga Program Among Children With Autism Spectrum Disorders: A Pretest–Posttest Control Group Design. *American Journal of Occupational Therapy, 66*(5), 538-546. doi: 10.5014/ajot.2012.004390

La Malfa, G., Lassi, S., Bertelli, M., Salvini, R., & Placidi, G. F. (2004). Autism and intellectual disability: a study of prevalence on a sample of the Italian population. *Journal of Intellectual Disability Research, 48*(3), 262-267. doi: 10.1111/j.1365-2788.2003.00567.x

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. doi: http://dx.doi.org/10.1016/0010-0285(74)90015-2

Lancioni, G. E., Singh, N. N., Oliva, D., Scalini, L., & Groeneweg, J. (2003). Microswitch clusters to enhance non-spastic response schemes with students with multiple disabilities. *Disability and Rehabilitation, 25*(6), 301-304. doi: doi:10.1080/0963828021000031179

Lang, R., Koegel, L. K., Ashbaugh, K., Regester, A., Ence, W., & Smith, W. (2010). Physical exercise and individuals with autism spectrum disorders: A systematic review. *Research in Autism Spectrum Disorders, 4*(4), 565-576. doi: http://dx.doi.org/10.1016/j.rasd.2010.01.006

Lee, S., Odom, S. L., & Loftin, R. (2007). Social Engagement With Peers and Stereotypic Behavior of Children With Autism. *Journal of Positive Behavior Interventions, 9*(2), 67-79. doi: 10.1177/10983007070090020401

Leonard, H. C., & Hill, E. L. (2014). Review: The impact of motor development on typical and atypical social cognition and language: a systematic review. *Child and Adolescent Mental Health, 19*(3), 163-170. doi:10.1111/camh.12055

Licciardello, C. C., Harchik, A. E., & Luiselli, J. K. (2008). Social skills intervention for children with autism during interactive play at a public elementary school. *Education and*

Treatment of Children, 31(1), 27-37.
Lord, C., & McGee, J. P. (2001). *Educating children with autism*: National Academy Press Washington, DC.
Lotan, M., Schenker, R., Wine, J., & Downs, J. (2012). The conductive environment enhances gross motor function of girls with Rett syndrome. A pilot study. *Developmental Neurorehabilitation, 15*(1), 19-25. doi: doi:10.3109/17518423.2011.629374
Maione, L., & Mirenda, P. (2006). Effects of Video Modeling and Video Feedback on Peer-Directed Social Language Skills of a Child With Autism. *Journal of Positive Behavior Interventions, 8*(2), 106-118. doi: 10.1177/10983007060080020201
Matson, J. L., Dempsey, T., & Fodstad, J. C. (2009). The effect of Autism Spectrum Disorders on adaptive independent living skills in adults with severe intellectual disability. *Research in Developmental Disabilities, 30*(6), 1203-1211. doi: http://dx.doi.org/10.1016/j.ridd.2009.04.001
Matson, J. L., Dempsey, T., LoVullo, S. V., & Wilkins, J. (2008). The effects of intellectual functioning on the range of core symptoms of autism spectrum disorders. *Research in Developmental Disabilities, 29*(4), 341-350. doi: http://dx.doi.org/10.1016/j.ridd.2007.06.006
Matson, J. L., & Kozlowski, A. M. (2011). The increasing prevalence of autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1), 418-425. doi: http://dx.doi.org/10.1016/j.rasd.2010.06.004
McGovern, C. W., & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *Journal of Child Psychology and Psychiatry, 46*(4), 401-408. doi: 10.1111/j.1469-7610.2004.00361.x
McLaren, L. (2007). Socioeconomic Status and Obesity. *Epidemiologic Reviews, 29*(1), 29-48. doi: 10.1093/epirev/mxm001
McNeill, A. W., & Mulholland Jr, R. (2010). Consolidation memory theory applied to relearning motor skills in profoundly retarded, multiply handicapped children. *APAQ, 3*(4).
Miyahara, M. (2013). Meta review of systematic and meta analytic reviews on movement differences, effect of movement based interventions, and the underlying neural mechanisms in autism spectrum disorder. *Frontiers in Integrative Neuroscience, 7*, 16. doi:10.3389/fnint.2013.00016
Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials, 16*(1), 62-73. doi: http://dx.doi.org/10.1016/0197-2456(94)00031-W
Moore, D., Aveyard, P., Connock, M., Wang, D., Fry-Smith, A., & Barton, P. (2009). *Effectiveness and safety of nicotine replacement therapy assisted reduction to stop smoking: systematic review and meta-analysis* (Vol. 338).
Mortimer, R., Privopoulos, M., & Kumar, S. (2014). The effectiveness of hydrotherapy in the treatment of social and behavioral aspects of children with autism spectrum disorders: a systematic review. *Journal of Multidisciplinary Healthcare, 7*, 93-104. doi:10.2147/JMDH.S55345
Muhle, R., Trentacoste, S. V., & Rapin, I. (2004). The Genetics of Autism. *Pediatrics, 113*(5), e472-e486.
Müller, R.-A., Shih, P., Keehn, B., Deyoe, J. R., Leyden, K. M., & Shukla, D. K. (2011). Underconnected, but How? A Survey of Functional Connectivity MRI Studies in Autism Spectrum Disorders. *Cerebral Cortex, 21*(10), 2233-2243. doi:10.1093/cercor/bhq296
Nash, C., & Collins, D. (2006). Tacit Knowledge in Expert Coaching: Science or Art? *Quest, 58*(4), 465-477. doi: 10.1080/00336297.2006.10491894
Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology/Revue canadienne de*

psychologie, 19(2), 155-160. doi: 10.1037/h0082899
Nickl-Jockschat, T., Habel, U., Maria Michel, T., Manning, J., Laird, A. R., Fox, P. T., . . . Eickhoff, S. B. (2012). Brain structure anomalies in autism spectrum disorder—a meta-analysis of VBM studies using anatomic likelihood estimation. *Human Brain Mapping, 33*(6), 1470-1489. doi:10.1002/hbm.21299
Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522-536. doi: 10.1037/h0026699
O Connor, J., French, R., & Henderson, H. (2000). Use of physical activity to improve behavior of children with autism-two for one benefits. *PALAESTRA-MACOMB ILLINOIS-, 16*(3), 22-29.
Owen-DeSchryver, J. S., Carr, E. G., Cale, S. I., & Blakeley-Smith, A. (2008). Promoting Social Interactions Between Students With Autism Spectrum Disorders and Their Peers in Inclusive School Settings. *Focus on Autism and Other Developmental Disabilities, 23*(1), 15-28. doi: 10.1177/1088357608314370
Pan, C. Y. (2011). The efficacy of an aquatic program on physical fitness and aquatic skills in children with and without autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1), 657-665.
Pan, C. Y. (2010). Effects of water exercise swimming program on aquatic skills and social behaviors in children with autism spectrum disorders. *Autism, 14*(1), 9-28. doi: 10.1177/1362361309339496
Peeters, T., & Gillberg, C. (1999). *Autism: Medical and educational aspects*: Wiley.
Petrus, C., Adamson, S. R., Block, L., Einarson, S. J., Sharifnejad, M., & Harris, S. R. (2008). Effects of Exercise Interventions on Stereotypic Behaviours in Children with Autism Spectrum Disorder. *Physiotherapy Canada, 60*(2), 134-145. doi:doi:10.3138/physio.60.2.134
Philip, R. C. M., Dauvermann, M. R., Whalley, H. C., Baynham, K., Lawrie, S. M., & Stanfield, A. C. (2012). A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neuroscience & Biobehavioral Reviews, 36*(2), 901-942. doi:http://dx.doi.org/10.1016/j.neubiorev.2011.10.008
Pitetti, K. H., Rendoff, A. D., Grover, T., & Beets, M. W. (2007). The efficacy of a 9-month treadmill walking program on the exercise capacity and weight reduction for adolescents with severe autism. *Journal of Autism and Developmental Disorders, 37*(6), 997-1006. doi: DOI 10.1007/s10803-006-0238-3
Pizzamiglio, M. R., Nasti, M., Piccardi, L., Zotti, A., Vitturini, C., Spitoni, G., . . . Morelli, D. (2008). Sensory-Motor Rehabilitation in Rett Syndrome: A Case Report. *Focus on Autism and Other Developmental Disabilities, 23*(1), 49-62. doi: 10.1177/1088357607311446
Reid, G., & Collier, D. (2002). Motor behavior and the autism spectrum disorders-introduction. *PALAESTRA-MACOMB ILLINOIS-, 18*(4), 20-27.
Rorie, A. E., & Newsome, W. T. (2005). A general mechanism for decision-making in the human brain? *Trends in Cognitive Sciences, 9*(2), 41-43. doi: http://dx.doi.org/10.1016/j.tics.2004.12.007
Rotheram-Fuller, E., Kasari, C., Chamberlain, B., & Locke, J. (2010). Social involvement of children with autism spectrum disorders in elementary school classrooms. *Journal of Child Psychology and Psychiatry, 51*(11), 1227-1234. doi: 10.1111/j.1469-7610.2010.02289.x
Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. (1972). A process model for long-term memory *Organization of memory* (pp. xiii, 423). Oxford, England: Academic Press.
Schulz, K., Altman, D., Moher, D., & Group, t. C. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(1), 18.
Shattuck, P. T. (2006). The Contribution of Diagnostic Substitution to the Growing Administrative Prevalence of Autism in US Special Education. *Pediatrics, 117*(4), 1028-

1037. doi: 10.1542/peds.2005-1516
Shin, I. S., & Park, E. Y. (2012). Meta-analysis of the effect of exercise programs for individuals with intellectual disabilities. *Research in Developmental Disabilities, 33*(6), 1937-1947. doi: http://dx.doi.org/10.1016/j.ridd.2012.05.019
Silver, W. G., & Rapin, I. (2012). Neurobiological Basis of Autism. *Pediatric Clinics of North America, 59*(1), 45-61. doi: http://dx.doi.org/10.1016/j.pcl.2011.10.010
Smith, N. J., Lounsbery, M. A., & McKenzie, T. L. (2014). Physical activity in high school physical education: impact of lesson context and class gender composition. *Journal of Physical Activity and Health, 11*, 127-135.
Sorensen, C., & Zarrett, N. (2014). Benefits of Physical Activity for Adolescents with Autism Spectrum Disorders: A Comprehensive Review. *Review Journal of Autism and Developmental Disorders, 1*(4), 344-353. doi: 10.1007/s40489-014-0027-4
Sowa, M., & Meulenbroek, R. (2012). Effects of physical exercise on Autism Spectrum Disorders: A meta-analysis. *Research in Autism Spectrum Disorders, 6*(1), 46-57. doi: http://dx.doi.org/10.1016/j.rasd.2011.09.001
Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., & Doll, E. A. (1984). *Vineland adaptive behavior scales*: American Guidance Service Circle Pines, MN.
Srinivasan, S. M., Pescatello, L. S., & Bhat, A. N. (2014). Current Perspectives on Physical Activity and Exercise Recommendations for Children and Adolescents With Autism Spectrum Disorders. *Physical Therapy*. doi: 10.2522/ptj.20130157
Stanfield, A. C., McIntosh, A. M., Spencer, M. D., Philip, R., Gaur, S., & Lawrie, S. M. (2008). Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies. *European Psychiatry, 23*(4), 289-299. doi:http://dx.doi.org/10.1016/j.eurpsy.2007.05.006
Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., . . . Schultz, R. T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *Journal of Child Psychology and Psychiatry, 51*(8), 944-952. doi: 10.1111/j.1469-7610.2010.02258.x
Terre, L., Drabman, R. S., Meydrech, E. F., & Hsu, H. S. (1992). Relationship between peer status and health behaviors. *Adolescence, 27*(107), 595-602.
Thiemann, K. S., & Goldstein, H. (2004). Effects of Peer Training and Written Text Cueing on Social Communication of School-Age Children With Pervasive Developmental Disorder. *Journal of Speech, Language, and Hearing Research, 47*(1), 126-144. doi: 10.1044/1092-4388(2004/012)
Utay, J., & Utay, C. (2005). Improving Social Skills: A Training Presentation to Parents. *Education, 126*(2), 251-258.
Vernazza-Martin, S., Martin, N., Vernazza, A., Lepellec-Muller, A., Rufo, M., Massion, J., & Assaiante, C. (2005). Goal Directed Locomotion and Balance Control in Autistic Children. *Journal of Autism and Developmental Disorders, 35*(1), 91-102. doi: 10.1007/s10803-004-1037-3
Virues-Ortega, J., Julio, F. M., & Pastor-Barriuso, R. (2013). The TEACCH program for children and adults with autism: A meta-analysis of intervention studies. *Clinical Psychology Review, 33*(8), 940-953. doi:http://dx.doi.org/10.1016/j.cpr.2013.07.005
Weintraub, K. (2011). Autism counts. *Nature, 479*(7371), 22-24.
Wuang, Y. P., Wang, C. C., Huang, M. H., & Su, C. Y. (2010). The Effectiveness of Simulated Developmental Horse-Riding Program in Children With Autism. *Adapted Physical Activity Quarterly, 27*(2), 113-126.

APPENDIX

A. AMSTAR scale

Score	Theme	Item
0/1	Prior design provided	Was an a priori design provided?
0/1	Duplication and data extraction	Was there duplicate study selection and data extraction?
0/1	Search literature comprehensively	Was a comprehensive literature search performed?
0/1	Publication as inclusion criterion	Was the status of publication (i.e., grey literature) used as an inclusion criterion?
0/1	Study list complied	Was a list of studies (included and excluded) provided?
0/1	Study characteristics provided	Were the characteristics of the included studies provided?
0/1	Quality assessment	Was the scientific quality of the included studies assessed and documented?
0/1	Proper use of quality assessment	Was the scientific quality of the included studies used appropriate in formulating conclusions?
0/1	Proper use of combination of findings	Were the methods used to combine the findings of studies appropriate?
0/1	Publication bias assessed	Was the likelihood of publication bias assessed?
0/1	Conflict of interest included	Was the conflict of interest included?

B. Items in the data extraction form

General information

- Identification features of the study Author
- Article Title
- Source (e.g. journal; conference; year; volume; pages; country of origin; and etc.)
- Design of the Study

Specific information

- Population characteristics and care setting
- Target population (describe)
- Inclusion criteria
- Exclusion criteria
- Recruitment procedures used (Participation rates if available)

Characteristics of participants at intervention commencement

- Number of participants
- Average age
- Gender and respective number of participants
- Number of participants in each condition (including control condition)

Interventions

- Number of conditions (including control condition)
- Content of intervention
- Intervention site
- Duration of intervention
- Time of intervention
- Number of session for intervention
- Delivery mode of Intervention
- Program rundown

Analysis

- Statistical techniques used
- Attrition rate
- Adjustment for confounding

C. Jadad scale

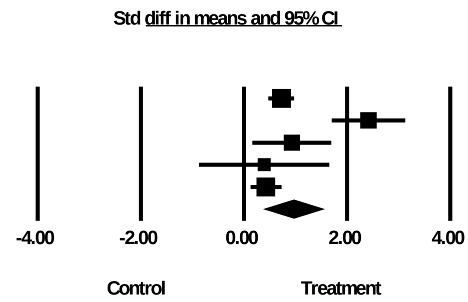
Score	Theme	Item
0/1	Randomization sounded	Was the study described as randomized (this includes words such as randomly, random, and randomization)?
0/1	Randomization procedure described	Was the method used to generate the sequence of randomization described and appropriate (table of random numbers, computer-generated, and etc.)?
0/1	Double-blinded	Was the study described as double blind?
0/1	Double-blinded procedure described	Was the method of double blinding described and appropriate (identical placebo, active placebo, dummy, and etc.)?
0/1	Withdrawals/dropouts described	Was there a description of withdrawals and dropouts?
0/-1	Inappropriate randomization process	Deduct one point if the method used to generate the sequence of randomization was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, and etc.)
0/-1	Inappropriate double-blinded process	Deduct one point if the study was described as double blind but the method of blinding was inappropriate (e.g. comparison of tablet vs. injection with no double dummy).

D. Modified CONSORT scale

Score	Theme	Item
0/1	Trial design	Description of trial design (such as parallel, factorial) including allocation ratio
0/1	Participants	Eligibility criteria for participants
0/1	Sample size	Greater than 5
0/1	Interventions	The interventions for each group with sufficient details to allow replication
0/1	Outcomes	Completely defined measures
0/1	Randomization	Method used to generate the random allocation sequence
0/1	Statistical methods	Statistical methods used to compare intervention and control groups for outcomes
0/1	Primary objective	Aiming at child with autism
0/1	Program dosage	Four weeks of intervention or above
0/1	Program variables	Sufficient variables to analyze

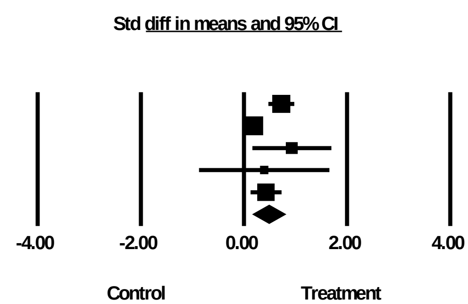
E. Supplementary section: analysis on the categories based on what type of exercises adopted

	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Aquatic exercise	0.725	0.122	0.015	0.485	0.964	5.929	0.000
Horse riding	2.415	0.358	0.128	1.713	3.117	6.742	0.000
Martial art	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Treadmill jogging	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Yoga	0.428	0.147	0.022	0.140	0.715	2.912	0.004
	0.969	0.282	0.080	0.416	1.523	3.433	0.001

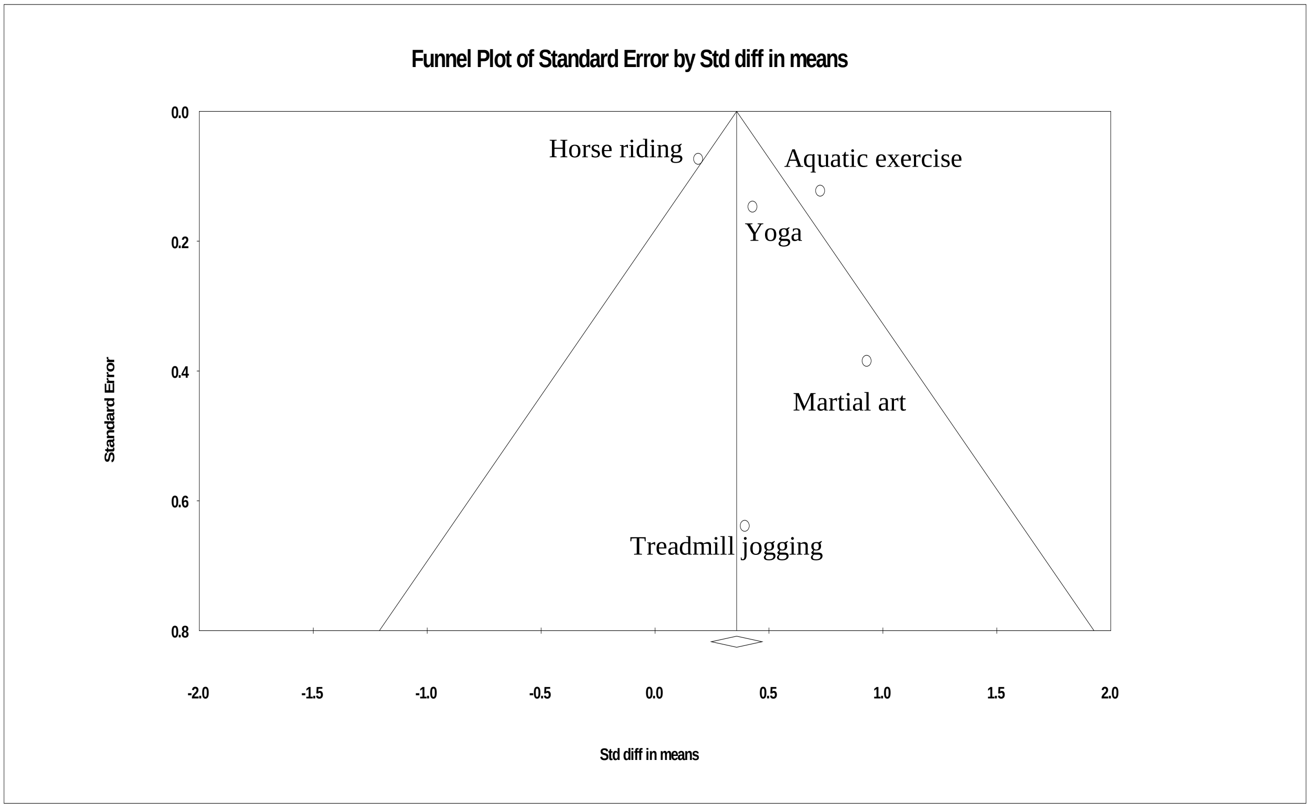


Effects of five exercise types were evaluated by its standardized difference in means (Cohen's d). The overall effect of all exercises was high (d = .969; 95% CI = .416, 1.523; p < .05; Q = 26.916, df (Q) = 4, p < .000). In that situation, horse riding was the one with highest effect size (2.415), however, since Wuang et al. (2010)'s study had detected an over-sensitive measurement scheme adopted, it should be excluded in the analysis its results would dominate the estimation of overall effectiveness. The re-calculation was conducted and the results were shown as below.

	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Aquatic exercise	0.725	0.122	0.015	0.485	0.964	5.929	0.000
Horse riding	0.189	0.073	0.005	0.046	0.332	2.587	0.010
Martial art	0.929	0.384	0.148	0.175	1.682	2.416	0.016
Treadmill jogging	0.394	0.639	0.408	-0.858	1.645	0.616	0.538
Yoga	0.428	0.147	0.022	0.140	0.715	2.912	0.004
	0.489	0.156	0.024	0.184	0.794	3.143	0.002



After excluding Wuang et al. (2010)'s study, the overall effect of all exercises was moderate to high (d = .489; 95% CI = .184, .794; p < .05; Q = 16.772, df (Q) = 4, p < .002)



After excluding Wuang et al. (2010)'s study, horse riding's effect size dropped sharply, from 2.415 to .189, and martial art became the one with highest score in effect size (.929) among all five exercises.

Given that the finding was different from the previous discussion [horse riding in Wuang (2010)'s study (effect size=4.529) was the one with highest effect, and after excluding it, the aquatic exercise in Pan (2010)'s study (effect size=1.593) became the one with highest effect]. However, another concern was raised out since only one study (Bahrami et al., 2012) was involved in calculating the effect of martial art, this study might not be that representative enough for all martial art interventions, so the final conclusion, still, could not be easily drawn.

CURRICULUM VITAE

Academic qualifications of the thesis author, SAM Ka Lam, Sam:

- Received the degree of Bachelor of Education (Secondary) (Honors) from The Hong Kong Institute of Education, November 2008.
- Received the degree of Master of Education in Physical Education (Distinction) from The Hong Kong Institute of Education, November 2009.
- Received the degree of Doctor of Education (Educational Measurement and Assessment) from The Hong Kong Institute of Education, November 2012.
- Received the degree of Master of Science in Exercise Science from The Chinese University of Hong Kong, November 2013.

December 2015