Fall 1-10-2020

A Case Study: Impact Of Assessment For Learning In A Badminton Unit In Physical Education

Lena Chng
Georgia State University

- ACCEPTANCE

This dissertation, A CASE STUDY: IMPACT OF ASSESSMENT FOR LEARNING IN A BADMINTON UNIT IN PHYSICAL EDUCATION, by LENA CHNG, was prepared under the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education and Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

_____
Jacalyn Lund, Ph.D.
Committee Chair

_____
Deborah Shapiro, Ph.D.
Committee Member

_____
Rachel Gurvitch, Ed.D.
Committee Member

_____
Michael Metzler, Ph.D.
Committee Member

_____
Date

_____
Jacalyn Lund, Ph.D.
Chairperson, Department of Kinesiology and Health

_____
Paul A. Alberto, Ph.D.
Dean, College of Education & Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with tis regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

_____

Lena S.H. Chng

NOTICE TO BORROWERS

All dissertation deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Lena S.H. Chng
Block 135, Serangoon Ave 3, #03-11
Singapore 556114

The director of this dissertation is:

Jacalyn Lund
Department of Kinesiology and Health
College of Education and Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Lena S.H. Chng

ADDRESS: 135, Serangoon ave 3. #03-11
Singapore 556114

EDUCATION:

Ph.D.	2019	Georgia State University, Kinesiology and Health – Physical Education Teacher Education
Masters Degree	2015	Georgia State University, Kinesiology and Health – Physical Education Teacher Education
Bachelor's Degree	1997	National University of Singapore, Arts and Social Sciences

PROFESSIONAL EXPERIENCE:

2018 to present	Senior Physical and Sports Officer (Physical Education) Ministry of Education, Singapore
2015 to 2018	Graduate Teacher Assistant Georgia State University
2011 – 2015	Senior Physical and Sports Officer (Physical Education) Ministry of Education, Singapore
2000 – 2011	Head of Department (Physical Education) Kuo Chuan Presbyterian Secondary School Ministry of Education, Singapore

PRESENTATIONS AND PUBLICATIONS

Chng, L. (2019) Formative assessment strategies for teaching of games and sports.

Singapore Teachers Conference 2019, May 27 – 29.

Chng, L. (2019) Assessment for Learning strategies in a Badminton Unit. Professional Learning

Zone organized by Physical Education and Sports Teachers Academy Singapore August 7.

Chng, L., & Gurvitch, R. (2018) Implementing plickers in health and physical education setting. *Journal of Physical Education, Recreation & Dance 89(2), 19-25,* DOI: 10.1080/07303084.2017.1404510

Chng, L & Lund, J. (2018) Assessment for learning in physical education: The what, why and how, *Journal of Physical Education, Recreation & Dance, 89:8, 29-34,* DOI: 10.1080/07303084.2018.1503119

Chng, L., & Lund, J. (2018) Impact of formative assessment in teaching of a badminton unit. *International Journal of Kinesiology in Higher Education*. https://doi.org/10.1080/24711616.2018.1509680

Chng, L. (2018) Impact of formative assessment in teaching of a badminton unit. *3rd FIEP Asia Conference on Physical Education and Sports, ACPES2018, 1-3 June.*

Chng, L. (April 2016) Implementing Plickers in health and physical education setting. Workshop presentation at Georgia State University Conference, Atlanta, GA.

Chng, L. (October 2015) Building 21st century skills through physical education. Poster Presentation at PETE/HETE Conference, Atlanta, GA.

Chng, L. (October 2015) Teaching and assessing creative dance. Workshop presentation at GAHPERD, Atlanta, GA.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2015-present Golden Key International, Member

2013-2016 S.H.A.P.E. America, Member

A CASE STUDY: IMPACT OF ASSESSMENT FOR LEARNING IN A BADMINTON UNIT IN PHYSICAL EDUCATION

by

LENA CHNG

Under the Direction of Jacalyn Lund

ABSTRACT

The main purpose of assessment for learning (AfL) is to provide feedback to learners, regarding the learners' progress towards the learning objectives, and allow students to take ownership of their own learning. The purpose of this study is to examine the impact of using an analytical rubric as an assessment for learning tool on teaching and learning in physical education. This case study examines the acquisition of skills in a badminton unit and student engagement in a Singapore's secondary school physical education class setting. Specifically, the research questions are: a) How does the incorporation of AfL tools affect students' acquisition of skills in a badminton unit? b) How does the incorporation of AfL tools in a badminton unit impact students' engagement (i.e., response rate)? A quasi-experimental group design method was used and one class had AfL tools incorporated in the teaching and learning, while the other class did not. Results showed that the class with AFL tools incorporated into the lessons had a significantly higher response rate than the class without. Both classes experienced almost similar improvement of skills even though the class with AfL tools incorporated had less game play time. This study concluded that assessment for

learning when incorporated into lessons can increase engagement and motivation, with no significant conclusion on the impact on skill improvement and psychomotor learning.

INDEX WORDS: assessment for learning, analytical rubrics, Singapore physical education

A CASE STUDY: THE IMPACT OF ASSESSMENT FOR LEARNING IN TEACHING AND LEARNING IN PHYSICAL EDUCATION

by

LENA CHNG

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Physical Education Teacher Education

in

Kinesiology and Health

in the

College of Education and Human Development

Georgia State University

Atlanta, GA
December, 2019

ACKNOWLEDGEMENTS

It has been an incredible four and a half years pursuing my Ph.D. after obtaining my Masters in Education with Georgia State University. This journey would not be not be possible without the support, guidance and encouragement of my mentor and friend, Dr. Jacalyn Lund. She has been the most patient with me, and her demand for excellence inspired me to put in my best efforts in all I do. My family and I are grateful to Dr. Lund and her husband Bill, who made sure our stay in Atlanta smooth and safe. I remember Bill having to drive three hours to another town with me, just so I could get my Georgia's Drivers License. We are thankful to blessed with such great friends. You both has been a huge part of our lives. Thank you Dr. Gurvitch, Dr. Shapiro and Dr. Metzler, for your advice and feedback. I am also thankful to the Ministry of Education (MOE), Singapore, for awarding my scholarship for my Masters Degree in GSU, without which I would not have stayed to pursue my Ph.D. The most important person I would like to thank is my supportive husband, Melvin Ong, whom sacrificed his career advancement, took a no-pay leave, just to move the family to Atlanta. I would not have finished my studies without him and his silent sacrifices. To my mom and dad, you both are my quiet heroes, supporting me throughout this journey. Thank you for making me who I am today. Not forgetting the Cubby Hole Kids: Gi-Choel who made my distance learning easier, you are always there when I needed you; Jay and Jin – the nicest couple who would help me with packing, moving, babysitting etc.; Dr. Andy Yao – thankful for your sharing and showing us the way to a Ph.D. despite English not being your first language. To my wonderful colleagues in PSOEB, thank you for making me look forward to going to work everyday. And finally, I devote the completion of this dissertation to my two boys, Evan and AJ. Thank you for your prayers and encouragement that your mom can do it! Your mom DID it! I love you all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. FORMATIVE ASSESSMENT IN PHYSICAL EDUCATION: A LITERATURE REVIEW

Formative assessment has been a central part of educational practice since the 1990s. Despite widening global awareness regarding the use of formative assessment, one review of assessment practices in U.S. schools stated that although formative assessment was widely promoted in the professional literature, it was not commonly used (Neil, 1997, p. 35-6). Clark (2008) suggested that perhaps because of this general lack of implementation that formative assessment was perceived by many as a new art of theory or classroom of practice.

Defining Formative Assessment

Scriven (1967) set the stage for the term formative assessment when he differentiated formative and summative evaluation. He wrote that formative evaluation was the process where information was gathered to assess the effectiveness of a curriculum and guide school system choices as to which curriculum to adopt and how to improve it (Scriven, 1967). Bloom, Hastings and Madaus (1971) used the formative evaluation concept to show how formative assessments could be linked to instructional units in a variety of content areas. In Bloom et al., (1971) summative assessment was used to grade students and/or evaluate the curriculum process. Formative assessment involved "the process of curriculum construction, teaching, and learning for the purpose of improving any of these three processes" (p.117). Bloom et al., (1971) determined that summative assessment is conducted at the end of a unit or curriculum for evaluation, while formative assessment is an ongoing process where the information gathered is used to make instructional decisions.

A popular distinction between formative and summative assessment by Sadler, (1989) follows:

> Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial and error learning. Summative assessment … is concerned with the summing up or summarizing the achievement status of a student, and is geared towards reporting at the end of a course of study especially for purposes of certification. (p.120)

According to Sadler, the focus of formative assessment was to improve student competence, while summative assessment was used to report the achievement of a student.

Black and Wiliam (1998) defined formative assessment as "all those activities undertaken by teachers and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities [and] when the evidence is actually used to adapt the teaching to meet students' needs" (p.2). In other words, assessment is formative if information about students' achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next step of instruction (by teachers) and learning (by students).

Table 1 shows some basic differences between summative and formative assessment used in classrooms. Although the distinction between formative and summative assessment lies in the purposes, many more practices are now associated with formative assessment, as it is now seen as embedded in classroom interaction between and amongst teachers and learners (Shepard, 2000). Teachers do not conduct tests only at the end of the unit, but include ongoing assessments to determine the progress of the students.

Assessment for Learning

The term formative assessment often was embedded in the term assessment for learning (AfL), following Gipps' (1994) distinction of AfL from assessment of learning (AoL). Although

many used the terms formative assessment and AfL synonymously, a few writers distinguished the two. The term "formative assessment" is used to describe the gathering of information that provides feedback to students about their own learning, and to inform teachers about subsequent planning and pedagogy (Black & Wiliam, 1998; Hay, 2006; Lund & Veal, 2013). Assessment is formative when the evidence is used to adapt teaching to meet learning needs (Black & Wiliam, 1998). Thus, formative assessment serves two purposes: (1) for teachers to review and adjust their teaching; and (2) for students to know about their progress. Assessment for learning is "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, what they need to do, and how best to get there" (Assessment Reform Group, 2002). In this definition, AfL is formative assessment in which learners play an active role in their own learning.

Table 1.1 Summary of Formative and Summative Assessment (adapted from Pregent, 2000)

	Summative Assessment	Formative Assessment
When	At the end of a unit / curriculum	Ongoing process during the unit
Purpose	For evaluation, to make a judgment	To improve learning
Feedback	Final judgment, no feedback to learner	Given to learner for information on learning progress
Frame of Reference	Normative (comparing students) or criterion	Criterion (evaluate students according to some form of rubric or criteria)
End product	Usually a grade or percentage	Descriptive feedback, no grades

While some writers equate the terms formative assessment and AfL, others make a distinction between the two terms. Formative assessment's main focus is to inform teachers of students' progress. The main purpose of AfL is to provide feedback to learners, regarding the learners' progress towards the learning objectives, and allow students to take ownership of their own learning. A well-planned AfL informs the most effective, meaningful and worthwhile

instructional strategies, to improve teaching and subsequently students' learning experience (Tannehill, Van der Mars, & MacPhail, 2013). MacPhail and Halbert (2010) explained the difference in this manner:

> Formative assessment is intended to enhance student learning through frequent opportunities for students to evidence their understanding, which in turn will identify ways to help individual students progress. While formative assessment tends to inform the teacher about student involvement, AfL (assessment for learning) extends to informing students about their own learning, acknowledging that they are decision-makers in their own learning. (p. 26-27)

Similarly, Chappuis, Commodore and Stiggins (2017), explained the difference as:

> Formative applications can diagnose student needs, monitor progress toward individual standards by individual students, and suggest changes in a teacher's instructional approach…. Assessment for learning includes those things that teachers and students do, in the classroom, to inform teaching and learning…. It helps students see and understand learning targets, helps them understand and manage their own progress, and uses self-assessment and goal setting to keep students connected to the targets of instruction. (p. 27)

The difference therefore, lies in the perspective and the use of formative assessment. Formative assessment becomes AfL, only when students are involved in the learning process, and they use the information to make decisions. Essentially AfL is a form of formative assessment, with the learners as its focus. While all assessments for learning are formative assessments, not all formative assessments are assessments for learning.

Types of Formative Assessment Activities

Many studies have tried to align formative assessment with contemporary psychological theories of learning. Early work on formative assessment focused on five main types of activities (Wiliam, 2000; Black, Harrison, Lee, Marshall, & Wiliam, 2003; and Wiliam, 2007):

1. Sharing success criteria with learners
2. Classroom questioning
3. Comment-only marking
4. Peer and self-assessment
5. Formative use of summative assessment

The following section will further explain these activities.

Sharing success criteria with learners. For learners to succeed, they need to know what success looks like. Teachers must inform students of the learning objectives and the criteria that they must reach. With intentions made clear, students will have a better understanding of what is expected of them, and will work towards that goal.

Classroom questioning. Teachers need to spend more effort in framing questions that are worth asking, i.e., questions that explore issues that are critical to the development of students' understanding. Students are not expected to answer the questions posed immediately, but are given adequate wait time, and time for every student to think and discuss the questions with their peers. The follow-up activities from the discussion must create opportunities for students to extend their understanding.

Comment-only marking. The provision of comments and feedback helps students focus on learning issues rather than trying to interpret a score or a grade given by the teacher. A numerical score or a grade does not tell students how to improve their work, and thus fails to enhance their learning. To improve learning, written assignments should encourage students to develop and

show understanding about the key features of what they have learned. The feedback comments by teachers should identify what was done well and what still needs improvement, and give guidance on how to make that improvement. Students should also have opportunities to respond to comments as part of the overall learning process.

Peer and self-assessment. The criteria for evaluating any learning achievements must be made transparent to students, so that they have a clear overview of both the aims of their work and what it means to complete it successfully. Students should also be taught the habits and skills of collaboration in peer assessment, and students should be encouraged to keep in mind the objectives of their work and how to assess their own progress. They have to be guided to become independent learners.

Formative use of summative assessment. Classroom practices can be improved by using summative tests for formative purposes. If students are engaged in a reflective review of the work they have done, they can plan revision(s) effectively. Students can also set their own questions and mark answers to gain an understanding of the assessment process and further refine their efforts for improvement. In addition, peer and self-assessments encourage students to apply criteria to help them understand how their work can be improved (i.e., providing opportunities for students to rework examination answers in class). In other words, summative assessments should become a positive part of the learning process, rather than just a judgment of how much students know. Summative assessment can be presented early as a form formative assessment in the learning process.

Creating a Theoretical Framework for Formative Assessment

While these five types of activities appeared to be connected to the central idea of formative assessment, there was no clear articulation of how they were connected. The lack of theoretical

foundation raised questions over if these five activities were representative of all the domains of formative assessment practices. In order to provide a better theoretical foundation for formation assessment, Wiliam and Thompson (2007) first drew on Ramaprasad's (1983), three key processes in learning and teaching:

- Establishing where the learners are;
- Establishing where they are going; and
- Establishing what needs to be done to get them there.

Crossing the three processes with the different agents (teacher, peer, learner), Wiliam and Thompson (2007) conceptualized a framework (Figure 1.1) indicating that formative assessment consists of five key strategies (Black & Wiliam, 2009):

1. Clarifying and sharing learning intentions and criteria for success;
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
3. Providing feedback that moves learners forward;
4. Activating students as instructional resource for one another; and
5. Activating students as owners of their own learning.

The framework that Wiliam and Thompson (2007), adopted used a historical cultural activity theory for the research analysis. They sought to unify the diverse set of practices described as formative. The framework they offered would potentially help open up new ways of helping teachers implement formative assessment more effectively.

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1. Clarifying learning intentions and criteria for success	2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3. Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4. Activating students as instructional resource for one another	
Learner	Understanding learning intentions and criteria for success	5. Activating students as owners of their own learning	

Figure 1.1 Aspects of Formative Assessment (Wiliam & Thompson, 2007)

Pryor and Crossouard (2008) used the social learning theory to conceptualize formative assessment, in the hope of engendering ongoing empirical research, particularly given the knowledge gap in current theoretical accounts. Formative assessment takes place when teachers and learners respond to student work, making judgment about what is good learning. Convergent and divergent formative assessment, adopted from Torrance and Pryor (1998), was used as another framework to discuss formative assessment. Convergent formative assessment starts from the goal of discovering if the learner knows, understands, or can do a predetermined thing. The teacher has a clear idea about what constitutes a correct response, and gives authoritative, judgmental or quantitative feedback on errors that contrast with the 'correct' response. The feedback focuses on the successful completion of tasks. Formative assessment is a means by which a teacher orchestrates the construction of a task, marking out a correct train of thought for students to complete the tasks (Pryor & Crossouard, 2008). Divergent formative assessment, on the other hand, seeks to discover what the learner knows, understands or can do. The questions teachers pose are different in that often, teachers themselves did not know the answer. Feedback is exploratory, provisional or provocative, prompting further engagement rather than correcting

mistakes (Pryor & Crossouard, 2008). The Divergent formative assessment is more aligned to the constructivist approach to learning. Divergent formative assessment, although is more student-centered, has the potential to open up social issues for all students, as not all learners come from the same social economic background. Students' experiences differ based on their social background thus making it difficult for teachers to build upon students' existing knowledge. Teachers' responsibilities are to encourage meta-social and meta-cognition reflection and discussion. Convergent and divergent formative assessment should not be seen as discrete categories, or as being good or bad. They should be seen as ideal types of formative assessments that could be placed at each end of the continuum (Torrance & Pryor, 1998).

Formative Assessment in Physical Education

Few scholars have studied how students are assessed in physical education classes. Veal (1998), in a study of 13 secondary physical education teachers' assessment practices, identified 90 assessment practices, of which 54% were summative assessments while only 30% were formative assessments. Imwold, Rider and Johnson (1982) conducted a survey on the method used to determine and report student performance in Florida public school physical education programs. They found that of the over 200 respondents, 18.4% did not use examinations of any type in their physical education program. Only slightly more than a half of the respondents used skills tests to assess their students. Less than 40% of the respondents used some form of written tests to evaluate their students (Imwold et al., 1982).

Despite many changes to assessment practice for most education during the 90s, there were few changes to assessment practices used in physical education during this time (Matanin & Tannehill, 1994). Matanin and Tannehill (1994) studied assessment practices in schools and found that there were few changes in assessment practices used in physical education in the 70s and 80s.

In physical education, assessment was tied to grading. Teachers used grades to hold students accountable for managerial issues such as effort and participation instead of performance and skills when grading. Grading was mainly tied to the physical fitness test performance, dressing and participation, attitude and behavior. Some teachers did administer written tests while standardized assessments of skills were rarely used (Matanin & Tannehill, 1994).

The climate of assessment in physical education seems to have moved forward in the 21st century, as there are more studies on the use of standards-based assessment, formative assessment, assessment for learning, and peer assessment. Michael and colleagues examined California physical education teachers use of standards-based assessment and grading practices (Michael, Webster, Patterson, Laguna, & Sherman, 2016). Standards-based curriculum refer to the skills, knowledge and dispositions that students should demonstrate to meet the standards set by the district, state or national committee (Lund & Tannehill, 2015). Standards-based assessment simply means assessing whether students achieved these standards. Michael et al., (2016) found that most teachers (n = 74.1%) reported using standards-based assessment and grading practices. Teachers who received professional development reported a greater use of standards-based assessment than teachers who reported not receiving the training (Michael et al., 2016). However, many teachers who used standards-based assessments also used non-standards-based assessments like administrative tasks, attitudes, and behavior to grade their students. The authors found that such practices might be due to large class size, lack of professional development for teachers, time constraints, and mixed grade level classes.

A review by Lopez-Pastor, Kirk, Lorente-Catalan, MacPhail, and Macdonald (2013), provided evidence of genuine progress in assessment in physical education, an area they felt has been fraught with difficulties as an educational endeavor. The traditional assessment instruments

such as physical fitness tests, and managerial behaviors such as grading students' effort and dressing for class were still widely used, but more physical education teachers are now using assessments that have a stronger educational focus. These new approaches included alternative assessment, authentic assessment, integrated assessment, learning-centered assessment, and assessment for learning (Lopez-Pastor et al., 2013). According to Siedentop and Tannehill (2000), alternative assessments are assessments that differ from the formal tools traditionally used in physical education, and instead involve students in actively solving realistic problems through application of new information, prior knowledge and relevant skills. Teachers felt that alternative assessment was a more relevant form of formative assessment when integrated in the teaching-learning process and when information about assessment was shared with the students (Desrosiers, Genet-Volet, & Godbout, 1997; Lopez-Pastor et al., 2013). When formative assessment was implemented in schools, there was evidence of improvement in learning, an increase of student involvement in the learning process, self-regulation, high reliability of students' self-assessment and self-grading, high student satisfaction and better grades (Lopez-Pastor, Monjas, & Manrique, 2011).

One commonly used form of formative assessment in research studies was peer assessment. Veal (1995) defined peer assessment as "when peers watch a partner performing a skill, and the peer is also learning about the skills, especially when the criteria have been clearly spelled out by the teacher" (p.14). Peer assessments, when implemented correctly resulted in more feedback for students and hence, improvement in learning, more sociability and more positive relationships among classmates (Butler & Hodge, 2001).

Formative assessment in physical education has a positive impact on student learning (MacPhail & Halbert, 2010). However, relatively few studies on the impact of formative

assessment in physical education exist (Hay, 2006). The next part of this paper will review the empirical studies about the use of formative assessments in physical education, and its impact on teaching and learning in physical education.

Method

Search Strategy

A search was performed in SPORTDiscus, ERIC, and Academic Research Complete using the terms: (a) “formative assessment AND physical education”; (b) “assessment for learning AND physical education”; and (c) “peer assessment AND physical education”. In addition, suitable articles that did not surface in the search but were cited in the articles reviewed were added.

Procedure

Six inclusion criteria were used to screen the articles. To be included in the review, articles had to be: (a) studying formative assessment or assessment for learning in physical education; (b) published between 1969, when formative assessment was introduced, and October 2017; (c) in general (not adapted) physical education setting; (d) written in English and published in scholarly (peer-reviewed) journals; (e) empirical studies; and (f) involving the K-12 (or equivalent age group) physical education setting. Reviews, book chapters, abstracts, posters, interviews, and narratives were excluded. Figure 1.2 shows how 22 articles were identified for this review.

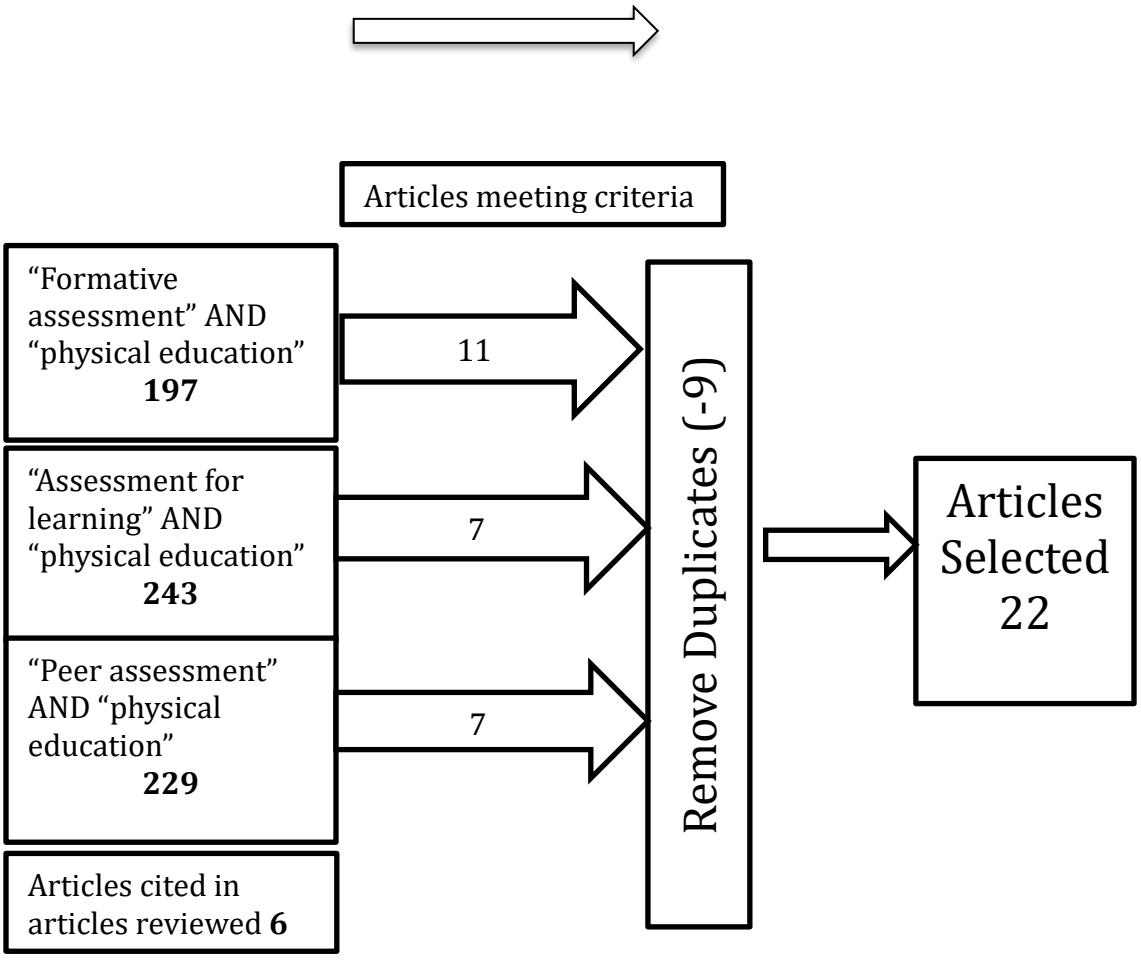


Figure 1.2 Article Selection Flow Chart

A total of 668 articles were screened from the database. After evaluating titles, abstracts and duplicates, 25 potentially relevant articles were retrieved. Six more articles were added after reviewing the articles. Eventually, nine articles were excluded as they were duplicates because the articles appeared in more than two of the search criteria. After assessing full texts, a total of 22 articles fulfilled all the inclusion criteria and were included in this review.

Results

Appendix A shows a list of the 22 articles chosen for this review. Most articles were from the United States (*n*=11), followed by Ireland, Norway and Spain contributing two articles each, and one article each from Australia, Germany, Greece, Hong Kong, and Sweden. Most articles used quantitative methods (*n*=9), and qualitative methods (*n*=6) design. Four articles used single-case study, while three used mixed method design. Most articles (*n*=11) studied primary students aged (5 to 13), while five articles studied secondary and high school students (ages 14 to 18). Six articles focused on teachers through questionnaire, interviews, case studies or focus groups.

After analyzing the 22 articles, they were grouped into categories of studies based on their purposes of study. It was found that nine articles looked at the impact of formative assessment (Chroinin & Cosgrave, 2013; Crouch, Ward & Patrick, 1997; Haug & Fischer, 2015; Johnson & Ward, 2001; Kolovelonis & Goudas, 2012; MacPhail & Halbert, 2010; Veal & Compagnone, 1995; Ward, Crouch & Patrick, 1998a; and Ward, Smith, Makasci & Crouch, 1998b), four articles concerned validation of some formative assessments tools designed (Hastie, Brock, Mowling & Eiler, 2012; Otero-Saborido & Gonzalez-Jurado, 2015; Otero-Saborido, Lluch & Gonzales-Jurado, 2015; and Richard, Godbout, Tousignant & Grehaigne, 1998), six studies researched teachers' perspectives on formative assessment (Brink & Bartz, 2017; Chroinin & Cosgrave, 2013; MacPhail & Halbert, 2010; Mintah, 2003; Penney, Jones, Newhouse & Campbell, 2012; and Yan

& Cheng, 2015), four articles studied students' perceptions on formative assessment (Butler & Hodge, 2001; Haug & Fischer, 2015; Penny et al., 2012; and Redelius & Hay, 2012), and four studies researched the implementation process of formative assessment (Hill & Miller, 1997; Leirhaug & Annerstedt, 2016; Leirhaug & MacPhail, 2015; and Mintah, 2003). Some articles appeared in two categories.

Based on the categories above, five themes emerged: (a) implementation of formative assessment; (b) validation of formative assessment tools; (c) impact of formative assessment; (d) teachers' perception of formative assessment; and (e) students' perception of formative assessment. The next section will provide an overview of the research involving each of these themes.

Implementation of Formative Assessment

This section explores how formative assessments or assessment for learning were implemented in schools, where the focus of physical education was argued to be 'more activity than learning' (Annerstedt, 2008).

In Norway, Leirhaug and MacPhail (2015), conducted a study to explore how formative assessment or assessment for learning (AfL) was understood and implemented by physical education teachers and the extent to which such implementation complements or challenges learning movement cultures within physical education. They selected and followed three teachers out of a focus group of 23 physical education teachers from a larger project, to study the perceptions of teachers and students on assessment in physical education (Leirhaug & Annerstedt, 2016). The teachers studied appeared to be well-informed about the national curriculum, assessment regulations and the purposes of AfL. However, their understanding and implementation of AfL was somewhat constricted. Researchers also observed that no real effort

was made to create the supportive environment in physical education for successful implementation of AfL strategies. Leirhaug and Annerstedt's (2016), study on teachers' and students' perspectives on AfL in physical education, showed that PE teachers conveyed varied understandings and implementation of AfL. That study noted some difference between teacher and student perspectives regarding AfL key principles, specifically that of the teacher providing feedback that moves learners forward. More studies are needed on implementation of formative assessment in physical education in schools.

Peer assessment. Peer assessment seemed to be the most common type of formative assessment used. Of the 23 articles reviewed, peer assessments were used in 14 articles (Butler & Hodge, 2001; Chroinin & Cosgrave, 2013; Crouch at el., 1997; Hill & Miller, 1997; Johnson & Ward, 2001; Kolovelonis & Goudas, 2012; Mintah, 2003; Otero-Saborido & Gonzalez-Jurado, 2015; Otero-Sabarido et sl., 2015, Richard et al., 1998; Veal & Compagnone, 1995; Ward et al., 1998a; and Ward et al., 1998b). Peer assessments can improve learning of skills (Chroinin & Cosgrave, 2013; Crouch et al., 1997; Johnson & Ward, 2001; Ward et al., 1998a; and Ward et al., 1998b). Several authors also determined that in order for peer assessment to be valid, students need to be trained to assess (Hill & Miller, 1997; Kolovelonis & Goudas, 2012; Otero-Saborido & Gonzalez-Jurado, 2015; Otero-Sabarido et al. 2015; and Richard et al., 1998). Hill and Miller (1997) had success with their students performing peer assessment on fitness testing. They reported that fifth grade students performed peer assessments on fitness performance with accuracy. They stated that their findings appear to support the growing body of research indicating that given proper training, educators in various disciplines can confidently utilize upper elementary students to record the test performance of their peers.

Age. Age of students is another variable in studies using formative assessment. Hastie et al., (2012), examined third graders' ability to accurately self-assess their performances on three dribbling tasks. Twenty third-graders were selected to perform some dribbling tasks and immediately following each task they assessed their performance through either live or video recall on four skill cues using a specific checklist. The authors found that students struggled with accurately assessing the dribbling cues. They were more accurately able to assess performance from video recall than live recall. In addition, boys were more accurate in assessing their own skill performance than girls but students of both sexes tended to overestimate than underestimate. Overall, the results of the Hastie et al., (2012) study indicated that the students in that study were minimally successful at assessing their own skill performance during one episode of dribbling, regardless of the type of recall.

Kolovelonis and Goudas (2012) had more success with fifth and sixth grade students. Their study showed that students were moderately accurate in self and peer recording, with the tendency to overestimate their performance. They also found that students who received more accurate feedback outperformed on the chest pass test over those who received less accurate feedback. In another study, fourth graders showed remarkable agreement in the use of a formative assessment tool for invasion games (Otero-Saborido & Gonzalez-Jurado, 2015). Therefore, it was concluded that upper elementary students can effectively perform peer assessments, given proper training.

Validation of Formative Assessment Tools

Four of the 22 articles looked at validating some form of formative assessment tool (Hastie, Brock, Mowling & Eiler, 2012; Otero-Saborido & Gonzalez-Jurado, 2015; Otero-Saborido, Lluch & Gonzales-Jurado, 2015; and Richard, Godbout, Tousignant & Grehaigne,

1998). Two studies were conducted to validate Team Sport Assessment Procedure (TSAP), developed by Grehaigner et al., (1997) as an assessment tool for team sport. Otero-Saborido et al., (2015) applied TSAP procedure as a formative assessment of invasion sport to 62 sixth graders in Spain and found a positive correlation between teacher observers' and students' assessments. The positive results reinforced the possibility of using this tool as a method of assessment in physical education. Richard et al., (1999) described the pilot of the TSAP in 6 elementary and junior high school physical education classes (grades five to eight) to identify potential issues related to its implementation. Teachers' reaction towards the integrating and usefulness of TSAP were very positive and informative. However, they added that students must be trained to use the tool, and reminded of possible variables. Overall, it was a good formative assessment tool for tactical games.

In summary, research indicated that when using peer and self-assessment, it is better for students to be trained, and reach a certain level of maturity before they can experience the benefits of peer assessment. The review above recommends that teachers implement peer assessment that is easy to use, and only for students at upper elementary and above level.

Impact of Formative Assessment on Student Learning

When formative assessment was introduced in physical education instruction, students performed better in terms of skills (Chroinin & Cosgrave, 2013; Crouch et al., 1997; Johnson & Ward, 2001; and Ward et al., 1998a) and had higher opportunities to respond (Ward et al., 1998b). Crouch et al., (1997), and Ward et al., (1998a), introduced peer-mediated instruction in an elementary physical education class, to determine if there was an effect on the number of hits and number of successful hits in a one-minute volleyball practice trials. Three different conditions: group instruction, peer-dyads, and peer-mediated accountability, were compared

using single-case method. In the group instruction condition, students progressed through the circuit of stations completing the one-minute trials. In the peer-dyads condition, students formed pairs to complete the one-minute trials. Students did not record their partners' performance, but were told to encourage their partners verbally. In the peer-mediated accountability condition, students in the same pair as the peer-dyads phase, kept a record of how many trials were correctly performed by their partners during the one-minute timing. Both studies found that students performed more trials and were generally more successful in the peer-mediated accountability conditions than the other two conditions. Peer-mediated accountability was also found to be effective in increasing opportunities to respond for both average and low-skilled students (Ward et al., 1998b). Researchers found that peer-mediated accountability was effective in promoting opportunities to respond, but inappropriate when students cannot perform the skill. The higher opportunities to respond may be due to students being more engaged in learning, when formative assessment was implemented during lessons (MacPhail & Halbert, 2010).

The inclusion of assessment in physical education lessons provided structure and focus to the planning, and thus has a positive impact on both teaching and learning (Chroinin & Cosgrave, 2013). Learners got immediate feedback on their performance and thus they were more engaged and more reflective, resulting in higher opportunities to respond and better performance as shown above. However, one study did find inconclusive results in determining the impact on students' self-perception of effort and skills (Veal & Compagnone, 1995). In that study, 151 sixth-graders participated in a series of teacher-designed formative assessments in basketball, volleyball, jump rope and badminton units. The types of formative assessment used in this study were teacher observation, self- and peer- assessment. The methods used in this study include students self-reporting successful passes each day during a basketball unit, with the

teacher posting results on a chart; students counting the number of successful hits over the net during badminton unit; teacher checking off students on her list of skills for jump rope; and students' individual recording of their badminton serves. The researchers felt that because assessment plans implemented in this study were not connected to any grades, students perceived that they were not important. They concluded that formative assessment needed to be linked to what will eventually be used for summative assessment.

Formative assessment is typically not recorded for a grade. However, the formative assessments eventually can be considered in the summative grade. The issue of the impact of accountability with formative assessment was explored in Haug and Fischer's (2015) study. Researchers eliminated the traditional grading in the physical education classes, and introduced a developmental continuum to 320 students, between 9 – 13 years of age. Researchers found that students showed far more intrinsic motivation in their learning when grades were removed. They took greater ownership over their development and showed more accountability for their own learning. However, this study did not show any data to support their findings (except that they reported 1% of the students preferred the traditional grades), nor did they mention the survey instrument they used. On the contrary, students in the Swedish secondary schools felt that grades were important (Redelius & Hay, 2012). Future research should look into whether students perform better when grades are removed, or whether students are still motivated by grades and the impact of accountability on formative assessment.

In sum, research showed that formative assessment did have an impact on teaching and learning. Teachers were better able to plan and structure their lessons, and students benefitted by being more engaged in the lessons, thus increasing response rates and eventually their success rates. Accountability seems to enhance the effectiveness of formative assessment. Formative

assessment would have to eventually link to summative assessment or some type of accountability to have a positive impact. Accountability played a huge role in deciding if formative assessment had an impact on teaching and learning.

Teachers' Perception of Formative Assessment

The inclusion of assessment in physical education lessons provided structure and focus to the planning, teaching and learning processes and impacted positively on both teacher and students (Chroinin & Cosgrave, 2013). The assessment strategies allowed for feedback related to assessment criteria and informed future planning for teachers (Chroinin & Cosgrave, 2013). Brink and Bartz (2017) examined high school teachers' perceptions and the use of formative assessment to enhance their planning, individualization of instruction, and adjustment of course content to improve student learning. Three teachers were studied over two years in a high school. Researchers found that when provided with specific information about formative assessment through staff development, teachers became more positive toward such assessment, and their implementation skills were greatly improved. Staff development had a positive impact on teachers as they were better equipped to differentiate instruction (Brink & Bartz, 2017). This study's results paralleled the findings in Yan and Cheng's (2015), study, which explored the relationships among teachers' attitudes, intentions, and practices regarding formative assessment under the framework of the Theory of Planned Behavior. Their results showed that although the theory of Planned Behavior was not an effective predictor of whether teachers used formative assessment, instrumental attitude, subjective norm, and self-efficacy were significant predictors of teachers' intention to use formative assessment in the future. Researchers from these studies concluded that if we want teachers to incorporate formative assessments into their instruction, they needed to be well-equipped with the skills to do so. Teachers with the skills to incorporate

formative assessments into their lessons have better planning, easier management and organization of lesson, felt motivated and energized (MacPhail & Halbert, 2010).

Mintah (2003) surveyed 210 public school physical education teachers to study the use of formative assessments in schools, and teachers' perceptions about the impact of formative assessment on students' self-concept, motivation and skill achievement. Formative assessments were used extensively in public schools (Mintah, 2003). In that study, 75.2% of physical education teachers used some form of formative assessment. The most popular techniques were teacher observation, self-observation, checklists, peer observation, and event tasks. The least popular types of assessments were portfolio and essay writing (Mintah, 2003). Because physical education is mostly about moving, many students did not like to write during physical education lessons. Physical education teachers must balance the use of formative assessment to enhance teaching and learning, while keeping students active. The study also showed that physical education teachers perceived that formative assessment had a positive impact on students' self-concept, motivation and skill achievement (Mintah, 2003).

Teachers in general had positive views towards the use of formative assessments in their instruction. Most importantly, teachers perceived that students are more engaged in lessons and learn better when formative assessments were introduced. Teachers are more inclined to incorporate formative assessments into their instruction if they are equipped with the skills to do so. Future research should continue to provide evidence of the positive impact of formative assessments in physical education.

Students' Perception of Formative Assessment

Formative assessment can enhance both accountability and student-centered learning (Butler & Hodge, 2001). Butler & Hodge (2001) conducted a case study on 24 high school

students during a softball unit, where peer assessment was introduced. Although only 20.8% of the students received specific written feedback from their peers, 91.7% of students felt that it was important to provide some form of feedback. The same number of students also felt that developing trust in someone is important for learning to occur (Butler & Hodge, 2001). Thus, students perceived receiving feedback and developing trust to be important, in peer (formative) assessment.

When no grades were tied to the activity, students showed more intrinsic motivation in their learning (Haug & Fischer, 2015). However, students in Veal and Compagnone's (1995) study showed the opposite. That study failed to show that formative assessment had an impact on students' self-perception of effort and skill. So, what do students really think about formative assessment? The answer may lie in whether students think the assessments were meaningful, and how the assessments were carried out.

Digital assessments appear to capture the attention of today's children. Western Australia tested out a digital form of assessment to 72 students in 4 schools, to determine if it could be implemented statewide (Penney et al., 2012). The digital forms include video-recording and computer-based capture of work, and computer-based marking. Participant observation and interviews showed that students perceived the assessment tasks to be authentic and meaningful. Students liked how the 'practical' and theoretical aspects were combined in the tasks (Penney et al., 2012). Digital assessments may be a useful tool to use when implementing formative assessments in schools.

In sum, peer assessments enhance trust among peers, and increase communication and socialization among students. Digital assessments may engage the students better when used as a formative assessment tool when some form of accountability is present. Students perceive a

greater benefit with formative assessments.

Discussion

Well-planned assessments, be they formative assessment, AfL, or peer assessment, inform the most effective, meaningful and worthwhile instruction strategies to improve teaching and learning. Formative assessments have a positive impact in physical education, as students perform better in terms of skills (Crouch et al., 1997; Holt et al., 2012; Johnson & Ward, 2001; and Ward et al., 1998a). The inclusion of formative assessment in physical education lessons provided structure and focus, thus helping teachers deliver lessons better (Chroinin & Cosgrave, 2013; MacPhail & Halbert, 2010). For formative assessments to be effectively implemented, both teachers and students needed to be equipped with the right skills. Students need to be trained to assist in their roles to peer assess and give feedback (Houston-Wilson et al., 1997). It seems that students as young as nine years old can be trained to perform formative assessment effectively (Otero-Saborido & Gonzalez-Jurado, 2015). Digital assessments appeared to capture students' attention in this current world overflowing with a variety of high-tech tools, from computer to video games, to sophisticated mobile devices that can record skill performances. More research should look into the use of digital tools for formative assessment in physical education in K-12 school settings.

Teachers favor incorporating formative assessment into their instruction, if they have knowledge and skills to do so (Mintah, 2003; Yan & Cheng, 2015). Therefore, staff development is important to expand teachers' understanding of formative assessment and the skills to implement it (Brink & Bartz, 2017). While Mintah (2003), found that formative assessments were widely used in the United States, there was little research about the extent of use and how these assessments were implemented. Leirhaug and Annerstedt's (2016) study showed that

physical education teachers conveyed varied levels of understanding and implementation of AfL. That study revealed some difference between teacher and student perspectives regarding AfL key principles, in particular regarding feedback that moves learners forward, which is a key component of formative assessment. In addition, it was noted that there were few research studies on other formative assessment practices other than the use of peer assessments. Future study should look at how various types of formative assessments are implemented in schools and how the data gathered are used.

Teachers perceived that students were more engaged in learning and were more reflective when formative assessments were implemented (MacPhail, & Halbert, 2010). Students showed more intrinsic motivation in their learning when no grades were tied to the activity (Haug & Fischer, 2015). However, students in Veal and Compagnone's (1995) study showed that formative assessment has no impact on students' self-perception of effort and skill. The researchers felt that this lack of impact was caused by the disassociation to grades and concluded that formative assessments need to be linked to knowledge and skills that they will eventually use for summative assessments. Given this contradiction, future studies are needed to examine whether students are more motivated when they are held accountable for a summative grade, or more motivated when formative assessment with no grades is used.

Conclusion

Formative assessment in physical education has undergone considerable development since Black and Wiliam's (1998) call to raise standards through classroom assessment. However, there has been limited research on impact of formative assessment (in general) in physical education (Hay, 2006), and the extent to which formative assessment was implemented. In addition, peer assessment is not the only form of formative assessment. There should be more

studies on other forms of formative assessment used. There is a need for more research in physical education on implementation of formative assessment, AfL, and/or other types of formative assessments other than peer assessment. Further exploration is requisite so that the profession can begin to track the extent to which these assessments were implemented as an integrated part of classroom practice. Assessment must be part of the instructional process to ensure quality teaching and learning, and ultimately move physical education forward.

References

Annerstedt, C. (2008). Physical education in scandinavia with a focus on sweden: a comparative perspective. *Physical Education and Sport Pedagogy, 13,* 303–318. doi:10.1080/17408980802353347

Assessment Reform Group, 2002. (retrieved from: http://www.assessmentforlearning.edu.au/default.asp)

Black, P., & Wiliam. D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan 80(2),* 139–44.

Black, P., & Wiliam, D. (2009) Developing the theory of formative assessment. *Educational Assessment, Evaluation & Accountability, 21*, 5-31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham: Open University Press.

Bloom, B., Hastings, J. & Madaus, G. (Eds) (1971) *Handbook on the formative and summative evaluation of student learning*. New York: McGraw-Hill.

Brink, M., & Bartz., D., (2017). Effective use of formative assessment by high school teachers. *Practical Assessment Research & Evaluation, 22(8),* 1-10.

Butler, S. & Hodge, S. (2001) Enhancing student trust through peer assessment in physical education, *The Physical Educator, 58(1),* 30-39.

Clark, I. (2008). Assessment is for learning: formative assessment and positive learning interactions. *Florida Journal of Educational Administration & Policy, 2,* 1-16.

Chappuis, S., Commodore, C., & Stiggins, R. (2017) *Balanced Assessment Systems: Leadership, quality and the role of classroom assessment*. Thousand Oaks, CA: Sage Publications.

Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical

education: teachers' perspective and experiences. *Physical Education and Sport Pedagogy, 18(2),* 219-233.

Crouch, D., Ward, P., & Patrick, C. (1997) Effects of peer-mediated accountability on task accomplishment during volleyball drills in elementary physical education. *Journal of Teaching in Physical Education, 17,* 26-39.

Desrosiers, P., Genet-Volet, Y. & Godbout, P. (1997) Teachers' assessment practices viewed through the instruments used in physical education classes, *Journal of Teaching in Physical Education, 16*(2*),* 211-228.

Gipps, C. (1994) *Beyond testing: Towards a theory of educational assessment.* London, Falmer Press.

Grehaigne, J., Bouthier, D., & Godbout, P. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education, 16(4),* 500-516.

Hastie, P., Brock, S., Mowling, C., & Eiler, K. (2012). Third grade students' self assessment of basketball dribbling task. *Journal of Physical Education and Sport, 12(4),* 427-430.

Haug, A., & Fischer, M. (2015) Improving teaching and learning in physical education: Using a developmental continuum and eliminating traditional grading. *SPORTSKE NAUKE I ZDRAVLJE 5(2),* 97-104.

Hay, P. (2006). Assessment for learning in physical education. In *The handbook of physical education,* ed. D. Kirk, D. Macdonald and M. O'Sullivan, 312-25. London: Sage.

Hill, G. & Miller, T. (1997). A comparison of peer and teacher assessment of students' physical fitness performance. *Physical Educator, 54,* 40-46.

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

Johnson, M., & Ward, P. (2001). Effects of class-wide peer tutoring on correct performance of striking skills in 3rd grade physical education. *Journal of Teaching in Physical Education, 20,* 247-263.

Kolovelonis A., & Goudas, M. (2012) Students' recording accuracy in the reciprocal and the self-check teaching styles in physical education. *Educational Research and Evaluation, 18(8),* 733-747, DOI: 10.1080/13803611.2012.724938.

Leirhaug, P. & Annerstedt, C. (2016) Assessing with new eyes? Assessment for learning in norwegian physical education. *Physical Education and Sport Pedagogy, 21(6),* 616-631.

Leirhaug, P., & MacPhail, A., (2015). It's the other assessment that is the key: three Norwegian physical education teachers' engagement (or not) with assessment for learning. *Sport, Education and Society, 20(5),* 624-640, DOI: 10.1080/13573322.2014.975113.

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

Lopez-Pastor, V., Monjas, R. & Manrique, J. C. (2011) Fifteen years of action-research as professional development: Searching more collaborative, useful and democratic systems for teachers. *Educational Action-Research, 19*(2), 153-170.

Lund, J., & Tannehill, D. (2016). *Standards-based physical education curriculum development*. Burlington, MA: Jones & Barlett Learning.

Lund, J. & Veal, M. (2013). *Assessment-driven instruction in physical education*. Champaign, IL: Human Kinetics.

MacPhail A., & Halbert, J. (2010) ‘We had to do intelligent thinking during recent PE’: students’ and teachers’ experiences of assessment for learning in post-primary physical education.

Assessment in Education: Principles, Policy & Practice, 17(1), 23-39, DOI: 10.1080/09695940903565412.

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

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

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

Neill, D. (1997). Transforming student assessment. *Phi Delta Kappan, 78,* 35-36.

Otero-Saborido, F., Lluch, A., & Gonzalez-Jurado, A. (2015). Student precision and reliability of team sport assessment in basketball: a primary education case study. *South African Journal for Research in Sport, Physical Education and Recreation, 37(2),* 83-94.

Oterio-Saborido & Gonzalez-Jurado (2015). Design and validation of a tool for formative assessment of invasion games. *Journal of Physical Education and Sport, 15(2),* 254-263.

Penney, D., Jones, A., Newhouse, P., & Campbell, A. (2012) Developing a digital assessment in senior secondary physical education. *Physical Education and Sport Pedagogy, 17(4),* 383-410.

Pregent, R. (2000) *Charting your course: how to prepare to teach more effectively*, Madison, WI: Atwood Publishing.

Pryor J., & Crossouard, B. (2008). A social-cultural theorization of formative assessment. *Oxford Review of Education, 34*: 1-20.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28,* 4–13.

Redelius, K., & Hay, P. (2012) Students views on criterion-referenced assessment and grading in physical education. *Physical Education and Sport Pedagogy, 17(2),* 211-225.

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

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science 18,* 119-144.

Scriven, M. (1967). The methodology of evaluation. In P. A. Taylor & D. M. Cowley (Eds.), Reading in curriculum evaluation (pp. 39–83). Dubuque, IA: Wm. C. Brown.

Shepard, L. (2000) The role of classroom assessment in teaching and learning, CSE Technical Report 517, Los Angeles, CA, National Center for Research on Evaluation, Standards, and Student Testing.

Siedentop, D., & Tannehill, D. (2000) *Developing teaching skills in physical education* (4th ed.) Mountain View, Mayfield.

Tannehill, D., Van der Mars, H., & MacPhail, A. (2013). *Building effective physical education programs*. Sudbury, MA: Jones & Bartlett.

Torrance, H., Pryor, J. (1998). *Investigating formative assessment*. Berkshire, England: Open University Press.

Veal, M. L (1995). Assessment as an instructional tool. *Strategies, 8,* 10-15.

Veal, M. L. (1988) Pupil assessment perceptions and practices, *Journal of Teaching in Physical Education, 7(4),* 327-342.

Veal, M., & Compagnone, N., (1995). How sixth graders perceive effort and skills. *Journal of*

Teaching in Physical Education, 14, 431-444.

Ward, P., Crouch, D., and Patrick, C. (1998a) Effects of peer-mediated accountability on opportunities to respond and correct skill performance by elementary school children in physical education. *Journal of Behavior Education, 8(1*), 103-114.

Ward, P., Smith, S., Makasci, K., & Crouch, D. (1998b). Differential effects of peer mediated accountability in elementary physical education. *Journal of Teaching in Physical Education, 17,* 470-478.

Wiliam, D. (2000). Formative assessment in mathematics part 3: the learner's role. *Equals: Mathematics and Special Educational Needs, 6(1),* 19–22.

Wiliam, D. (2007). *Content then process: Teacher learning communities in the service of formative assessment*. In D. Reeves (Ed.), Ahead of the curve: The power of assessment to transform teaching and learning (pp. 182–204). Bloomington, IN: Solution Tree.

Wiliam, D., & Thompson, M. (2007). *Integrating assessment with instruction: What will it take to make it work?* In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 53–82). Mahwah, NJ: Erlbaum.

Yan, Z. & Cheng, E. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teacher and Teacher Education, 45,* 128-136.

2. A CASE STUDY: ASSESSMENT FOR LEARNING IN A BADMINTON UNIT IN PHYSICAL EDUCATION

Formative assessment and assessment for learning (AfL) in physical education has undergone considerable development since Black and Wiliam's (1998) call to raise standards through classroom assessment. Formative assessments have a positive impact in physical education, as students perform better in terms of skills (Crouch, Ward, & Patrick, 1997; Holt, Kinchin, & Clark, 2012; Johnson & Ward, 2001; and Ward, Crouch, & Patrick, 1998a). The inclusion of formative assessment in physical education lessons provides structure and focus, thus helping teachers deliver lessons more effectively (Chroinin & Cosgrave, 2013; MacPhail & Halbert, 2010). Students tend to be more on-task, and have a higher response rate when formative assessment is incorporated into lessons (Chng & Lund, 2018). Teachers and students need the proper skills and knowledge to implement formative assessments effectively. Training is necessary for students to peer assess and give feedback (Houston-Wilson, Dunn, Van der Mars, & McCubbin, 1997). Students as young as nine years old can be trained to perform formative assessment effectively (Otero-Saborido & Gonzalez-Jurado, 2015).

Historically, physical education teachers have administered few assessments and typically use assessments summatively for grading (Veal, 1988). Teachers would incorporate formative assessment and AfL into their instructions if they have knowledge and skills to do so (Mintah, 2003; Yan & Cheng, 2015). Therefore, staff development is important to expand teachers' understanding of formative assessment and the skills to implement it (Brink & Bartz, 2017; Van der Mars, McNamee & Timken, 2018). Leirhaug and Annerstedt's (2016) study showed that physical education teachers conveyed varied levels of understanding and implementation of AfL. That study showed limited implementation of AfL principles in physical

education, especially assessments that involve students in the learning process, and providing meaningful feedback for students to move forward (Leirhaug & Annerstedt, 2016). When the feedback component is missing, formative assessments may not result in significant changes in students' learning (Chng & Lund, 2018). The importance of professional development was further reinforced in a study by Van der Mars and colleagues (2018). In their study, the combination of professional development workshops, on-site coaching, and prompting enabled teachers to increase their competency for incorporating formal-formative assessment of students' performance. They reinforced the need for professional development to be ongoing, and long-term (Van der Mars et al., 2018).

There has been limited research on the impact of formative assessment (in general) in physical education (Hay, 2006), and the extent to which formative assessment and AfL was implemented (Mintah, 2003). The proposed study seeks to further the research on implementation of formative assessment, assessment for learning, and/or other types of formative assessments and their impact on teaching and learning in physical education.

The Singapore Context

Singapore has always prided itself on its place in the education world. Singapore's education system has been examination-oriented (Cheah 1998; Gopinathan 2001; Lim-Ratnam 2013). Students' advancement to the next level is determined solely by examinations taken at the end of each level (e.g., students need to pass the end of year examinations in Primary Five to advance to Primary Six; students need to pass end of year examinations in Secondary Two to advance to Secondary Three). An unintended consequence of the examination-oriented system is over-emphasis on grades and content acquisition, instead of a focus on learning and holistic development (Tan, 2016). In 2018, the Ministry of Education (MOE) Singapore, announced the

abolishment of some exams at certain levels, especially during the transition years (i.e., Primary One and Two, Secondary One and Three; MOE, 2018). MOE Singapore encouraged schools to move away from many summative examinations, to focus on a more formative approach to promote greater joy and intrinsic motivation in learning.

Physical education requirements in Singapore. The Singapore Ministry of Education rolled out the revised physical education syllabus in 2014 (MOE, 2013). The revised syllabus placed emphasis on equipping students with competencies to engage in a wide range of physical activities and sports. Students are expected to demonstrate the physical skills, practices, and the right attitude and values to enjoy a lifetime of active healthy living. (MOE, 2016) As such, the Secondary Physical Education Syllabus focused on developing students' skills in at least six physical activities and sports for active participation beyond school. All secondary students take part in outdoor education and additionally, participate in at least three recreational sports during intra-school competitions. Health education lessons are integrated into PE lessons at all levels.

Assessment in Physical Education. The Ministry of Education Singapore takes a balanced approach when it comes to assessment in physical education (MOE, 2016). MOE Singapore encouraged schools and teachers to provide a balanced perspective on students' progress and achievement in physical education. Although the MOE Singapore did not dictate what and how to assess, it did provide general guidelines on conducting assessment in the 2014 PE Syllabus. For instance, MOE Singapore encouraged schools to collect assessment data from multiple sources, occurring at frequent intervals and over a period of time, to determine whether students achieved the goal of physical education. The intent was to use assessment to provide continual feedback for students and promote self-monitoring and peer evaluation to encourage responsibility for one's own learning (MOE 2016, pp. 217 – 222).

Developing game rubrics. Rubrics identify criteria used to evaluate students' performances. They also have the potential to contribute to quality instruction. As Popham (1997) pointed out, rubrics can guide a teacher in designing lessons because it is students' mastery of the evaluative criteria that ultimately leads to skill mastery. Thus, the use of rubrics can be two-fold: to evaluate and report a student's performance (assessment *of* learning), and to guide teachers and students in their teaching and learning (assessment *for* learning).

MOE Singapore developed a set of analytical rubrics to aid teachers in planning and implementing learning progressions, and in assessing whether students achieved the learning outcomes of each sport. Rubrics were developed for the following nine sports: badminton, basketball, floorball, netball, soccer, ultimate, softball, table tennis and volleyball. To ensure content validity of this set of rubrics, the Physical Education Assessment Committee (PEAC) was tasked to determine the validity of the rubrics in physical education. The PEAC committee, chaired by the Physical, Sports and Outdoor Education Branch (PSOEB) of Ministry of Education Singapore, consisted of members from the following: Physical Education and Sports Teachers Academy (PESTA) who is responsible for in-service professional development of physical education teachers, Physical Education and Sports Science (PESS) who conducts training and certification of Pre-service teachers, school leaders (e.g. principals and vice principals) who were formerly heads of departments of physical education, and current heads of department of physical education. PEAC members and physical education teachers who are experts in the area of sports (i.e., coaches and former national players) were asked to provide feedback on the rubrics. PSOEB trialed and tested the validity and reliability for two of the games that they deemed most difficult to attain reliability – softball and ultimate.

The game rubrics were written to assess students' ability to play the culminating game as described in the 2014 PE Syllabus. The indicators written are based on the concepts of the game, and are similar across games of similar tactics. Four attainment levels (beginning, developing, achieved, and exceeding) describe performance of students in game situations. The 2014 PE Syllabus recommended at least 16 hours of instruction in order for students to achieve the games learning outcomes stated. For students to attain the learning outcomes of the game, they must reach the 'achieved' level of competence. Although rubrics were written to evaluate students' performance, they can also guide teachers when planning lesson progressions, and determine students' learning when teachers use them to provide feedback to students. Students are aware of their current level of performance according to criteria specified on the rubrics, understand the expected performance outcome (end goal), and set personal improvement goals. A copy of the Badminton game play rubric is shown in Appendix B.

Purpose of study

Assessments for learning have a positive impact in physical education, as students perform better in terms of skills (Crouch et al., 1997; Johnson & Ward, 2001; and Ward et al., 1998a). The studies cited were all conducted at the elementary level. The study by Chng and Lund (2018), looked at the impact of formative assessments on the response and success rate of eight students in a sixth-grade badminton physical education class. Although that study showed a higher success rate for the students who had formative assessments throughout the unit, the difference between the treatment and control class was not significant. Lack of statistical difference could be due to the low number of students in the study.

The purpose of the current study is to examine the impact of learning in PE when assessment for learning tools are incorporated into the teaching and learning of a badminton unit.

This study examines the acquisition of skills using an analytical badminton game rubric, and student engagement in a badminton unit. Specifically, the research questions were: a) How does the incorporation of AfL tools affect students' acquisition of skills in a badminton unit? b) How does the incorporation of AfL tools in a badminton unit impact students' engagement (i.e., response rate)?

Method

Participants and Settings

The study took place in a secondary school in Singapore. The school was selected based on the physical education teacher's experience in implementing assessments for learning (AfL) in physical education lessons. One experienced physical education teacher from a secondary school was invited to participate in this study. The teacher was a 39-year-old male, senior PE teacher, in an all-boys secondary school in Singapore, with over 11 years of teaching experience. The teacher worked with the researcher in incorporating assessment for learning tools into his physical education lessons and was willing to participate in this project. He was chosen because of his experience with implementing assessment for learning into his lessons. This teacher has shared during many professional development platforms in Singapore on how he incorporates AfL tools into his PE lessons.

Classes in Singapore's secondary schools are typically sorted by their academic results. Students are assigned to classes based on school-based examinations conducted at the end of each year. For this study, the teacher selected two classes that were similar in terms of academic abilities (i.e. normal academic stream) with class sizes of 30 and 19 respectively. The researcher explained the purpose of the study to the students of the two selected classes and issued assent

and parental consent forms to the students. The forms were collected by the teacher. For Class A, 20 out of 30 students agreed to participate in the study. All 19 students in Class B agreed to the study. However, due to poor attendance, three students from Class B were dropped from the study. The study took place during the physical education lessons of the two selected classes. The lessons were conducted in the Indoor Sports Hall (ISH) of the school, which houses four badminton courts. The study was conducted in the months of April to May 2019, during the physical education lessons, which were on Tuesdays and Fridays, from 8am to 9am for Class A, and 9am to 10am for Class B (one hour, twice a week). Due to the design of the intervention (i.e., treatment class had AfL strategies such as peer- and self-assessment incorporated in the lessons while control class did not), the first class with 30 students was assigned the treatment class, while the second class with 16 students was designated as the control class. With AfL tools integrated into the lessons, the teacher could engage all 30 students in the ISH at the same time, as half the class were peer-assessing when they were not playing. For this reason, the class with 30 students was the treatment class. Students in the study put on colored numbered bibs for easy identification by the researcher. Students not in the study were not given a numbered bib. Students not in the study were placed at the last court where no video cameras were set up. Non-study students were told that should they be accidentally captured in the video, the researcher would not code their performance and there was no way the researcher could identify them.

The researcher and teacher planned 10 lessons for the study. However, due to other commitments of the school (i.e., Founders' Day, and Public Holiday), only 8 lessons were taught and recorded. Details of the unit and assessment plans are shown in Appendix C.

A badminton unit was chosen for this study because badminton skills are discrete and easier to quantify (i.e., successful or unsuccessful shots) as compared to other sports with more

continuous skills like running in space and dribbling. Students' responses can be quantified by counting the number of times the students hit the shuttle, and whether the response was a successful hit during the game situation. The Ministry of Education Singapore's badminton rubric (Appendix B) was used to determine the skill levels of the students. As badminton is a complex game (i.e., it can be played cooperatively and competitively), the number of successful and unsuccessful shots alone is unable to determine the skill level of the students. As the rubric was written developmentally, it is able to capture the different performance level of the students during a game play situation. Measurement of skill performance in a game play setting makes the evaluation of the performance more valid. Although PSOEB did not test the reliability of the rubrics for badminton, the researcher felt confident using the badminton rubric to report the skills progression of the students, as the content, format and descriptors for the rubric were validated by experts in the field of badminton. PE teachers who are also badminton coaches, validated the content of the rubric. In this study, inter-observer agreement was conducted between the two independent raters who rated the students on their skills progression using the rubrics. The two independent raters were officers from PSOEB, and were trained to use the rubrics. These two raters were involved in the design of and validation of the rubrics. They have been in consultation with PE teachers and trialed the use of the rubrics in schools.

Materials

Permission was obtained from the Ministry of Education Singapore to collect data from the school. After obtaining IRB approval for the research study, the researcher contacted the school and teacher to discuss lesson design and data collection. The researcher and the teacher agreed on the classes chosen for the study. Students' assent and consent from parents of the students in the classes selected was sought. Students who did not have signed assent and consent

forms for the study were not video recorded, but they continued to take part in the lessons. Students in the Class A who did not give consent in the study were placed at the last court where no cameras were set up. They were told that if they were accidentally filmed during the class, researchers would not be able to identify them as they were not wearing a numbered bib.

The researcher worked with the teacher to use the badminton analytical rubric (Appendix B) to plan the lessons, and as a self- or peer- assessment for learning tool throughout the unit. For example, the teacher used logs for students to self-record their successful attempts at certain types of shots to determine if students could execute those shots, and used heat maps for students to understand their performance and whether they hit the shuttle away from opponents. If a student successfully executed at least two different types of shots, the student was rated 'developing' under the return shot indicator (i.e., hit the shuttle over the net using two different shots) using the rubric. The heat map was used to capture the placement of shots for every player, allowed teachers and students to know if they have reached the 'achieving' level (i.e., hit the shuttle into open space, away from the opponent, using a variety of shots). If the heat map showed shots that were near the nets, it was assumed that students could execute a drop shot or a net shot. If there were shots at the back of the court, it showed that students could execute an overhead clear or an underhand lift. The teacher used the information on the heat map to determine if students were able to execute a variety of shots and planned his lesson progression accordingly. For instance, after observing that not many students hit the shuttle to the front of the court, near the net, in the following lesson the teacher placed more emphasis on net shots and drop shots.

The heat map was used in lessons six and seven. After peers recorded the performance of their partners on the heat map, the players would analyze their own heat map to understand the

areas on the court for which they were successful in returning shots. Players would also look at their opponent's heat map, to analyze where they placed their shots. Players then reflected on their game performance and wrote down improvement goals for themselves. Students also did a self-assessment of their performance on the rubric, to know their level of attainment and identify areas on which they needed additional work. Such was the characteristic that differentiated formative assessment with assessment for learning, where students took ownership of their own learning and set performance goals for themselves. The teacher then collected the heat map at the end of the lesson, so that he could analyze the data, and wrote individual feedback on the heat map to the students. This feedback was given to the students the following lesson for them to read and understand, before the start of the activity. This process of students using the data (teacher feedback and peer assessment information) to set individual improvement goals, and the teacher using the data to make informed decision of the lesson progression was done throughout the lessons in the treatment class, through a variety of assessment tools. A copy of the unit and assessment plan is shown in Appendix C.

Design and Procedure

Content validity of the badminton rubric was established previously by the PEAC and PE teachers who are also badminton coaches for the Singapore Combined Schools Badminton team. Two independent physical education teachers, who are also PSOEB officers, were trained to use the rubrics to rate performance level of students at the beginning (Lesson 1) and at the end of the unit (Lesson 8) for both the treatment and control class. The two raters used the badminton rubric to rate performance levels of the students independently. As there were some disagreements after the first round of evaluation (25% disagreement), the two observers met to reconcile discrepancies and agreed on the performance level of students by re-watching the video

recording of the lessons together and rating the performances together. A copy of the final ratings from the two observers is shown in Appendix D.

Students were trained to self- and peer- assess during the lessons. The teacher explained the rubric to students, and showed videos of what each learning outcome meant prior to the start of the unit. During the first lesson, students were trained to peer-assess before they started to perform the peer assessment task. The teacher asked two students to play a rally, while the rest of the class recorded the number of successful and unsuccessful hits. The teacher explained if a hit was successful with every single shot. After this initial training, the two students played a one-minute rally, and the rest of the class recorded on their papers the number of successful and unsuccessful hits. The teacher then checked if the class recorded the correct number of successful and unsuccessful hits. The training was done in a similar fashion for every different peer-assessment task throughout the unit. This training helped ensure that the feedback students received on their performance was mostly accurate.

A quasi-experimental group design was adopted for this study to observe the impact of using assessment for learning tools, on students' skills acquisition, and engagement during the lesson. For this study, the physical educator taught a badminton unit for eight lessons of one hour per lesson over five weeks. The 2014 PE Syllabus recommended that schools should provide at least 16 hours of lessons for students to achieve the learning outcomes. The teacher planned 10 hours of instructions but could only carry out eight. The teacher planned to teach that another eight hours of badminton lessons the following year.

The teacher selected two classes of the same academic level to participate in this study. Class A served as the treatment class (i.e., AfL group) where AfL tools and strategies were incorporated into the teaching and learning. There were 30 students in Class A, but only 20

signed consent forms to allow participation in the study. The teacher introduced the game rubric to students at the beginning the unit and students used it as a self-assessment tool to determine their personal level of play. Students were aware of the learning outcomes expected of them. They were told that they should aim to reach at least at the 'Achieving' level in the badminton rubric. The rubric was used as an AfL tool, for students to reflect on their performance and set personal goals throughout the unit, using peer and self-assessments (Appendix C).

Other product assessment tools (e.g., counting the number of times students accurately served into the service box, mapping the players' returned shuttles successfully, and where players missed the return of shots) were used to prompt students when they reflected on their performances levels using the rubric throughout the unit. The badminton rubric was used again by the raters at the end of the eight lessons to measure the post-lesson performance of the students. Appendix C shows an example of a peer- and self- assessment tool that was used with the badminton rubric. The descriptors in the badminton rubric were used to develop self-assessments for students.

To establish accuracy of peer assessment, students' ability to peer-assess was checked by comparing their observation results with those of the researcher's. In the first lesson, students were asked to code with their partners successful and unsuccessful hits in a 10-minute badminton game. A successful hit was defined as one in which the student successfully hit the shuttle over the net and into the playing area during a game. An unsuccessful hit was defined as unable to hit the shuttle, or unable to return the shuttle into the play area. The researcher also coded the same game play by watching a video recording of the lesson, to assess the students' accuracy in coding. Inter-observer agreement was calculated based on the number of agreements, divided by the number of agreements and disagreements. The inter-observer agreement between the students

and the researcher was 85.3% (Appendix E). The researcher was confident that students were able to record the performances and give accurate feedback to their peers.

The second class, Class B, served as the control class where the teacher taught the badminton unit without any use of formative assessment or assessment for learning tools. The teacher used the same drills and teaching progressions as in the treatment class, except without any assessment. The only form of feedback students in this group received was through the teacher's observation and verbal feedback during the lessons. Besides having more activity time (as the students did not perform peer-assessment tasks), there was no difference in the practice tasks between the control and the treatment class.

To examine total learning time for both classes, the researcher recorded how time was spent during the lessons (i.e., management time, instruction time, activity time, etc.). Appendix F shows the time recorded for knowledge (i.e., teacher's instructions, demonstration, and teaching tasks) and activity (i.e., drills, practice, peer assessment and game playing) in minutes for both the control and treatment class. Total learning time is the combination of knowledge and activity time. From the graphical representation in Appendix F, it was shown that the activity minutes for the control class was slightly higher than the treatment class (Control = 253.6 minutes; Treatment = 213.7 minutes). Due to the need for instructions on peer assessment, the treatment class had higher knowledge time (Control = 80.6 minutes; Treatment = 152.1 minutes). However, the total learning time (sum of knowledge and activity time) was almost similar for both classes (Control = 334.2 minutes; Treatment = 365.8 minutes).

Data Analysis

The dependent variable in this study was the improvement of skill levels of each student as measured by the difference of skill level ratings on the badminton rubric between the first and

last lesson. Skill level improvement was measured by the extent to which the students of the class improved in levels (i.e., beginning, approaching, achieving and exceeding) using the descriptors specified on the game rubric. The skill rating of each student was evaluated by two independent physical education teachers trained to use the rubrics. The independent physical education teachers viewed the recorded performances of the first and the last lesson to assess each student's level of skill for each indicator listed on the badminton game play rubric (i.e., serve, return shots and space coverage). Students in both treatment and control classes were assessed. Analysis results are shown in Appendix D.

Additionally, the quality of students' responses was coded by successful or unsuccessful hits during game play for all lessons. A successful hit was defined as students successfully hitting the shuttle over the net and into the playing area during a game, while an unsuccessful hit was defined as unable to hit the shuttle, unable to return the shuttle into the playing area, or did not attempt to hit the shuttle that was in the court. Game time refers to the time students were physically and actively playing a game of badminton during the lesson. It does not include drills, instructional time, time spent recording during peer assessments tasks, time spent setting up and wait time etc. Data were analyzed by success rates of the students during game play. Success rate was calculated by number of successful hits divided by number of successful and unsuccessful hits, and multiplied by 100 to determine a percentage. Success rates for each lesson were tracked for individual students across the unit.

Students' engagement was measured by the students' response rate during game play in each lesson. Students' response rates were calculated by the total number of hits of the badminton shuttle (successes and misses) divided by the entire game play time of the lesson. The response rate is presented in number of hits per minute of game play time. A statistical test

(independent T test) using SPSS was conducted to determine if there were any significance differences between the means of the students between the two classes.

To further analyze the benefit of AfL tools to students of differing abilities, the students were categorized by their skill levels into More Skilled, Average Skilled and Less Skilled based on their performance during game play in the first lesson. Students who were More Skilled group were identified by having more than 70% successful hits during the game, and at least one area (i.e., Serve, Return Shots or Space Coverage) with a rating of 'Developing' in the rubrics. Students who were Less Skilled group were identified by having fewer than 60% successful hits during the game and with a 'beginning' rating for all areas. Those between the More and Less Skilled players were placed into the Average Skilled group. In the treatment class (n=20), there were eight students in the Higher Skilled group, six students in the Average Skilled group, and 6 students in the Less Skilled group. In the control class (n=16), there were six students in the Higher Skilled group, five students in the Average Skilled group, and five students in the Less Skilled group. Appendix G shows the raw data of students' performance (i.e., successful and unsuccessful hits) and amount of game play time (to calculate response rate per minute).

Validation of Findings

To ensure validity of findings, the content and pedagogical approaches must be similar amongst the two classes. The researcher was present for every lesson to ensure fidelity of content and instructional practices. The teacher used the same drills and teaching progressions in both the treatment and control classes. The only difference was that the treatment class had AfL tasks where they had to record their performances, while the control class did not. To ensure fidelity of the implementation of the AfL tools, the researcher planned with the teacher and observed every lesson, to ascertain that the teacher implemented the AfL tools as planned.

To determine inter-rater reliability of the students' performances, students were video-recorded during two assessment points (one in the beginning of the unit and the other at the end of the unit). Two trained raters independently evaluated the students' performances according to the rubric. The two raters were officers from PSOEB (i.e., trained PE teachers) who had trialed the use of the rubrics with schools. The initial rater agreement was 75%. The two raters met a second time to view the videos, focusing on the performances of the students with disagreements between the two independent raters. After watching the video and discussion, the two raters resolved the discrepancies and agreed on the skills ratings as shown in Appendix D.

To determine reliability of data coding of response rates, an independent trained observer coded 25% of the lessons for the treatment and control class (i.e., on Day 3 and Day 6). IOA was calculated by the total number of agreements divided by the total number of agreements and disagreements (Appendix H). The IOA calculated was 90.35%.

Results

Skills Improvement

The dependent variable used in this study was improvement of skills of each student as measured by two independent raters using the game rubric, on the first and last lesson. Skill improvement was measured by the extent to which students in the class improved in their levels (i.e. beginning, approaching, achieving and exceeding) using the indicators specified on the game rubric. Table 2.1 and Table 2.2 show the results of the skill level improvement for the control and treatment class respectively. In the control class, nine out of 16 (56%) students showed improvement in their skills levels. Two students *(HS)* showed improvement in all three areas (i.e., serve, return shots and space coverage), two students *(HS = 1; AS = 1)* showed improvement in two areas (i.e., return shots and space coverage), while six students *(HS = 3; AS*

= 2; LS = 1) showed improvement in one area (i.e., five in return shots, and one on serve). In the treatment class, nine out of 20 (45%) students showed improvement in their skills levels. Of these nine students, two students (*HS = 1; AS = 1*) showed improvement in all three areas (i.e., serve, return shots and space coverage), one student (*MS*) showed improvement in two areas (i.e., return shots and space coverage), while six students *(HS = 3; AS = 1; LS = 2)* showed improvement in one area (i.e., three in return shots, and three on space coverage).

Table 2.1 Skill Improvement for Control Class

SS Nos.	Group	Lesson 1			Lesson 8		
		Serve	Return Shots	Space Coverage	Serve	Return Shots	Space Coverage
4	Higher Skilled	1	1	1	1	2*	1
7		1	2	1	2*	3*	2*
12		1	2	1	1	2	1
14		1	2	1	1	3*	2*
15		1	2	1	2*	3*	2*
19		1	1	1	1	2*	1
1	Average Skilled	1	1	1	1	1	1
2		1	1	1	1	1	1
9		1	1	1	1	2*	2*
13		1	1	1	1	2*	1
16		1	1	1	1	2*	1
6	Less skilled	1	1	1	1	1	1
10		1	1	1	1	1	1
11		1	1	1	1	2*	1
17		1	1	1	1	1	1
18		1	1	1	1	1	1

** students who showed improvement*
 # 1= Beginning; 2 = Developing; 3 = Achieving

Table 2.2 Skill Improvement for Treatment Class

SS Nos.	Skill Group	Lesson 1			Lesson 8		
		Serve	Return Shots	Space Coverage	Serve	Return Shots	Space Coverage
2	More Skilled	1	2	1	1	2*	2*
4		1	2	1	2*	3*	2*
8		1	2	2	1	2	2
10		2	3	2	2	3	2
12		1	2	1	1	2	1
13		1	2	2	1	2	2
14		1	2	1	1	2	2*
15		1	2	1	1	2	2*
3	Average Skilled	1	1	1	1	2*	1
7		1	2	1	2*	3*	2*
11		1	2	1	1	2	1
16		1	1	1	1	2*	2*
18		1	1	1	1	1	1
19		1	1	1	1	1	1

1	Less Skilled	1	1	1	1	1	1
5		1	1	1	1	2*	1
6		1	1	1	1	1	1
9		1	1	1	1	2*	1
17		1	2	1	1	2	1
20		1	1	1	1	1	1

** students who showed improvement*
 # 1= Beginning; 2 = Developing; 3 = Achieving

Upon further analysis of the different skill level groups of students, out of the nine students who showed improvement in the treatment class, four were from the Higher Skilled group, three from the Average Skilled group, while two from the Less Skilled group. In the Control class, of the nine students who showed improvement in their skill levels, five were from the Higher Skilled group, three were from Average Skilled group, while only one was from the Less Skilled group.

Most of the improvements made were made on the 'Return Shots' and 'Space Coverage' indicators, and not many on 'Serve'. This is attributed to the teacher instructing students on one type of serve to start the game, and not covering high and low serves as indicated in the descriptions for 'Serve'. These different serves were meant to be taught at a later stage. Similarly, the majority of the unit focused on 'Return Shots' of the players, and only two lessons were on 'Space Coverage'. Thus, the most improvement was on 'Return Shots'. The rubric was written based on the 16 hours of teaching in a badminton unit. For the current study, only eight hours of instruction was completed which could explain why not many students were at the 'Achieved' level. With only eight hours of instructions, the rubric adequately captured that most students were at 'Developing' level which was considered appropriate by the Singapore MOE. The next eight hours of badminton instruction will occur the following year (outside the time

frame of the current study). The rubric is a developmental rubric that can be used for all skill levels, from beginning to advanced.

Success Rate

The quality of students' responses was coded by differentiating successful or unsuccessful hits during the game play for all lessons. Figure 2.1 shows the comparison of percentage of successful hits (i.e., total number of successful hits, divided by total number of successful and unsuccessful hits, multiplied by 100) of the treatment and control class. Students in the control class started a little below the students in the treatment class (Treatment = 67%, Control = 64%), and after 8 lessons, classes achieved success rates of 71% and 69% respectively.

An independent samples t-test was conducted to compare the percentage of successful hits in the treatment and control class, using SPSS. There was no significant difference in the mean scores for the treatment class (M=69.13, SD = 1.25) and the control class (M=67.13, SD=3.27); t(14)=1.616, p=.128.

Table 2.3 further breaks down the success rate of students by skill levels, making comparison between the first and eighth lesson. In the treatment class, the students in the Averaged Skilled and Less Skilled group showed the most improvement (AS = +5%; LS = +6%), while the Less Skilled students in the control class showed the most improvement (LS = +11%).

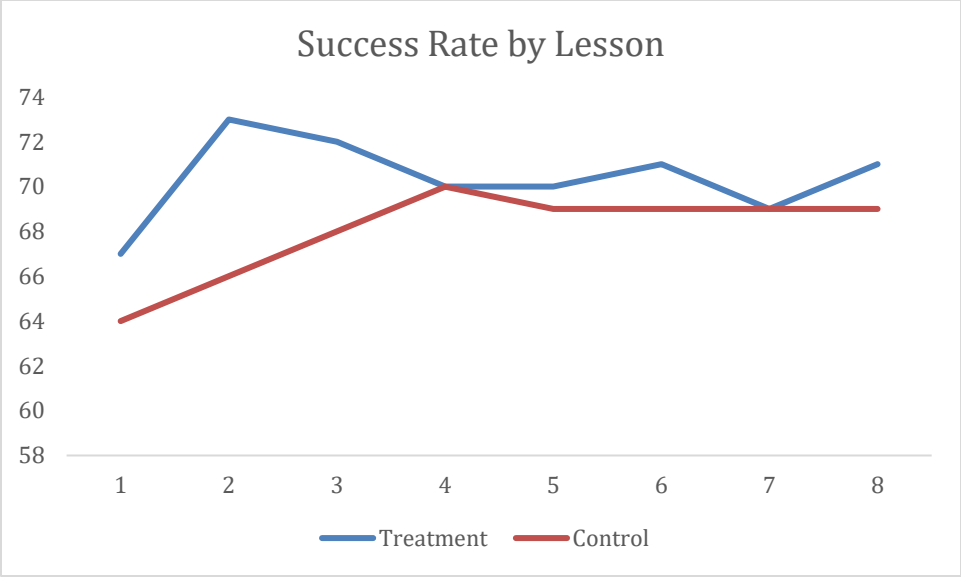


Figure 2.1 Comparison of Success Rate between Treatment and Control Class

Table 2.3 Comparison of Success Rate by Skill Levels

	Treatment			Control		
	Lesson 1	Lesson 8	Improvement	Lesson 1	Lesson 8	Improvement
More Skilled	75%	76%	+1%	75%	78%	+3%
Averaged Skilled	69%	74%	+5%	68%	68%	-
Less Skilled	54%	60%	+6%	45%	56%	+11%

Response per Minute

Students' response rates were calculated by the total number of hits of the badminton racquet and shuttle (successes and misses) divided by the time spent on game play during the lesson. The response rate is presented in number of hits per minute of game time. Figure 2.2 shows the comparison of the response rates per minute for both the treatment and control class. The treatment class had higher response rates per minute in every lesson, especially the last four lessons. An independent sample t-test was conducted to compare the response per minute for the

treatment and control classs. There was a significant difference in the scores for the treatment class (M=11.20, SD = 1.12) and the control class (M=7.55, SD=1.64); t(14)=5.204, p=.000.

Table 2.5 shows the amount of game play time between the treatment and control class and the average number of hits by the students in these two classes. The students in the control class had almost doubled the amount of game play time for lessons one, two, three, four and eight, and almost four times more game play time for lessons five, six and seven. Despite having more game play time, the number of hits by the students in the control class did not increase proportionately during the last four lessons. In fact, the treatment class has almost twice the response rate as the control class during the last four lessons.

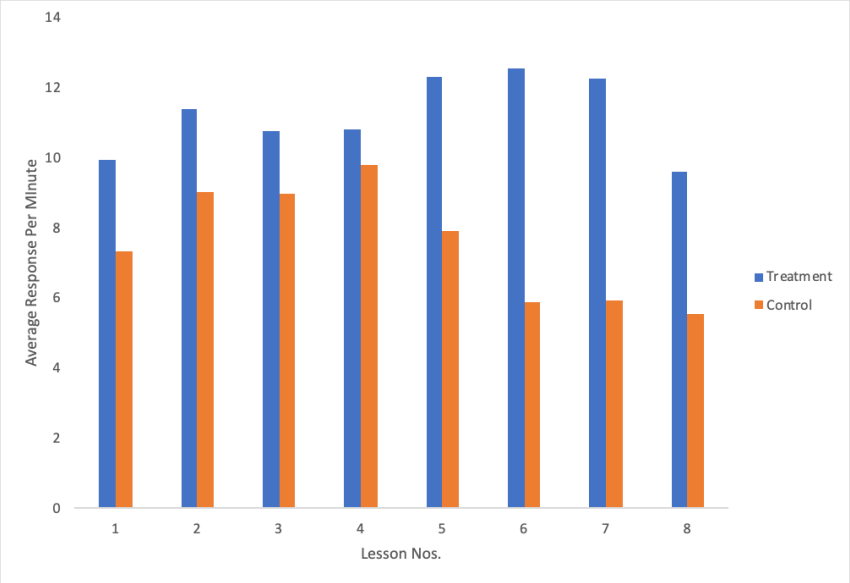


Figure 2.2 Comparison of Response Rate per Minute between Treatment and Control Class

Table 2.4 Comparison of Mean Response Rate per Minute by Skill Levels

Skill Levels	Treatment Class	Control Class
More Skilled	11.34	7.82
Average Skilled	10.86	7.44

Less Skilled	10.43	7.27

Table 2.5 Comparison of Game Play Time and Response Rate per Minute between Treatment and Control Class

	Lesson 1	Lesson 2	Lesson 3	Lesson 4	Lesson 5	Lesson 6	Lesson 7	Lesson 8
Treatment:								
Average Game Play Time per student	9.14	10.01	16.29	15.79	10.29	9.60	8.77	16.17
Average Hits per Student	90	115	175	171	127	121	107	155
Control:								
Average Game Play Time per student	28.00	20.53	23.95	30.92	38.50	23.00	34.50	29.83
Average Hits per Student	205	185	215	302	304	136	205	165

A further analysis was done on the mean response rate for students of different skill levels (Table 2.4). From Table 2.4, students from all skill levels in the treatment class where AfL tools and strategies were incorporated in the lessons, had higher mean response rates per minute than students in the control class. The group that had the highest mean response rates per minute was the More Skilled group, followed by the Average Skilled group, and the Less Skilled group for the treatment. For the control class, there was not much difference in terms of the response rate of students with different skill levels, ranging from 7.27 for the Less Skilled group to 7.82 for the More Skilled group.

Discussion

The research questions for this case study were: a) How does the incorporation of AfL tools affect students' acquisition of skills in a badminton unit? b) How does the incorporation of AfL tools in a badminton unit impact students' engagement (i.e., response rate)? In the current

study, students' acquisition of skills is determined by the success rate of the students, and the rating on the performance rubric.

There are few studies on the impact of formative assessment in physical education settings. In the studies identified, students performed better in terms of skills when formative assessment was incorporated into instruction (Crouch et al., 1997; Johnson & Ward, 2001; Ward et al., 1998a). These studies used a one-minute skills test for basketball and volleyball to measure improvement in skills. In the current study, which measured students' performance in a game setting, results were different. Students who performed well in closed skills test setting may not necessarily perform well in open games situation, where it is more dynamic and complex. The application of skills in game play is more difficult and improvement requires more time. The one minute skill tests were an artificial setting unlike the current study that studied student improvement in an authentic environment. In the current study, a badminton rubric was used to measure students' skills in an authentic game play context. As the skills test measurements are different (one in a closed skill context and another in an open skill games context), this may account for the differences in the outcomes. Also, students in those studies were elementary students, who were not as skilled to begin with. The current study's participants were in secondary school and the ages of students were between 13 to 14 years. There was only one other study that used intervention study on the impact of formative assessment in physical education in a middle school setting (Chng & Lund, 2018). No other studies were found measuring the impact of assessment for learning in an authentic game situation.

Using the performance rubric to show students' skill acquisition, Tables 1a and 1b showed that the treatment class started out slightly more skilled than the control class with more students in the 'Developing' level. The treatment class had 11 students with at least one indicator

at the developing level, while the control class had only 4. At the end of the eight lessons, the control class had slightly higher percentage of students who showed improvement in the skills levels (56%) as compared to the treatment class (45%). This could be due to the fact the students who started out skilled need more time and practice to show improvement, as compared to students at the beginning level.

In terms of success rate, the treatment class started out with a mean success rate of 67% and after eight lessons, the success rate was 71%. For the control class, they started out slightly lower at 64% and ended with 69% after eight lessons. A statistical analysis of the mean scores of the success rates did not result in a significant difference between the treatment and control class. This replicates the study by Chng and Lund (2018), which also saw no significant difference between the group with formative assessment, and the group that did not use formative assessment. Similarly, the group that saw the most improvement in succes rate were the Less Skilled students from the control class. This also replicated the findings from Chng and Lund (2018), where they discussed that lower skilled students need more practice time to improve their skills. Additionally, students in the lower skilled group tended to focus on cooperative play and maintaining a rally, which leads to higher success rates. Students with advanced skills hit the shuttle away from the opponent which leads to lower success rates. Therefore, using success rate alone is not a good representation of students' skill acquisition.

From Figure 2.1, it was obvious that the improvement in success rate was prominent in both class in the first four lessons, and reached a plateau after Lesson 5. Badminton is a complex game which can be played cooperatively (trying to maintain a rally with your opponent) and competitively (hitting away from the opponent to score a point). The teacher focused on cooperative play during the first four lessons and taught the the concept of hitting away from the

opponent from lesson 5 onwards. As the game got more competitive, the likelihood of returning the shuttle decreased. This was also the limitation cited in Chng and Lund's (2018) study. Although the control class had more game time than the treatment class, there was no significant difference in the success rates of the two groups. This could mean that having more quality instruction and assessment for learning, is as important as practice time during game play, where students do not constantly get feedback from a teacher. A game performance rubric was introduced in this current study as a means for examining the skill level improvement of students. Students in this study were somewhat skilled. If the study had lasted longer, one can hypothesize that more skill improvement would have occurred.

Further analysis on the different abilities of the students showed findings worth noting. Students in the More Skilled groups showed the greatest improvement in terms of the game performance rubric in the control class, while in the treatment class, students in all three groups showed improvement in the skill level rating. This could mean that AfL strategies benefit students of all abilities, as compared to just letting students play in the game with little feedback from teachers. In the study by Ward, Smith, Makasci, and Crouch (1998b), researchers found that peer-mediated accountability (peer assessment) was effective in promoting opportunities to respond for both averaged and low skilled students, but inappropriate when students cannot perform the skill. On surface, this study may seemed to contradict the findings, but as mentioned earlier, the students in this current study were considered fairly skilled when the study began. Additionally the test measurement was different (i.e., closed skill testing versus open games skills).

In this study, engagement level was measured by response rate (i.e., number of times students hit the shuttle) per minute. It was assumed that when students hit the shuttle, they are

engaging in some form of playing the game. During game time, students who were engaged would be actively hitting the shuttles and playing the game, while those who were not engaged would be engaging in off-task behaviour (i.e., not hitting the shuttle). In terms of engagement level, which was measured by the response rate per minute, it showed that the treatment class had a significantly higher response rate than the control class. This means that students in the treatment class were hitting the shuttles more during the time allocated to play the game. Although the control class had more game time, students were not as engaged (i.e., hitting the shuttle as many times) as those in the treatment class that used AfL. This is especially true in lessons six to eight where the average number of hits in the control class was only slightly higher than that of the treatment class despite having four times more game play time. The AfL activities kept the students focused in the treatment class whereas the control class was not. It is the focus of activities, and accountability through AfL tasks that made students more engaged during lessons.The treatment class had AfL tasks incorporated during some of the game time and their performances were recorded on the heat map, while no recording on the game performance was conducted for the control class. Students in the treatment class were held accountable for their learning through AfL activities. Students who were held accountable tended to be more engaged in their learning (Lund & Shanklin, 2011). This finding was similar to the that of MacPhail and Halbert's (2010), and Chng and Lund's (2018) study which attributed higher student response rates to being more engaged in learning when formative assessment and assessment for learning tasks were integrated in the lessons.

Similarly, peer-mediated accountability was found to be effective in increasing opportunities to respond for both average and low-skilled students (Ward et al., 1998a). Researchers found that peer-mediated accountability was effective in promoting opportunities to

respond, but inappropriate when students cannot perform the skill. This case study replicated those results. From Table 2.3, it can be seen that all students from all skill levels in the treatment class where AfL tools and strategies were incorporated in the lessons, had higher mean response rate per minute than the students in the control class. The group that had the highest mean response rate per minute was the More Skilled group, followed by the Average Skilled group, and the Less Skilled group. This confirmed the conclusion by Ward and colleagues (1998b), that AfL may be an effective way to promote opportunities to respond, but less so for students who were less skilled.

The students' engagement level could be also seen by the researcher who was present throughout the study. When students were given AfL tasks, they received feedback on their performances immediately. Also, they knew that the teacher would give written feedback on their performances (i.e., on their heat map) and hold students accountable for their performances. Students who were held accountable tended to perform better and were more on task (Lund & Shanklin, 2011). This may be the motivating factor for students staying on task as explained in Vroom's (1964) expectancy model where effort leads to desired outcomes. In the treatment class, every student's performance during game play was recorded by a peer and feedback on performance was immediate. In the Control class, during game play, students were either counting rallies or playing to a number of points to see who won the game. The only form of performance feedback they received was from the rally points or game points, or occasional feedback they received from the teacher. As there were 19 students in the class, the teacher could not give every individual student feedback on their performances at all times. Thus, the students on video were observed to be laughing a lot, rolling on the floor, hitting the shuttle at their friends instead of passing to them to start the serve etc. They took the game lightly as there was

no accountability. This is similar to Hastie's (1998) study on physical education teacher and students behavior. His study showed that where teachers are more passive and and did not hold students accountable, students tend to display more off-task behavior.

Therefore, although the length of practice time is important for students to improve on their skill acquisition, it is more important that practice time is purposeful. French, Rink, Rikard, Mays, Lynn, and Werner (1991), suggested that tasks should focus on a single skill or concept, and have progression in levels of difficulty. The treatment class in this study had a different task focus in every lesson, and students were held accountable through the use of AfL tools like peer assessment and self assessment. The control class on the other hand had more game time, but were not held accountable for their game practice. Although the treatment class had less game time, they had significantly higher response rate per minute than the control class.

Limitations

There were some limitations to this study. First, there were only eight lessons conducted, which means students had only eight hours of physical education lessons to learn badminton. In such a short timeframe, it is difficult to see much skill improvement in students learning such a complex game. Furthermore, badminton is a game played by many students in Singapore. Students would have learnt the basics of the game in the Primary School. Most students had fairly good skills to begin with. With only eight lessons in this study, the study may not yield much improvement in terms of skills. Second, the sample size was too small to generalize the impact of AfL in teaching and learning; Thus, this study was presented as a case study. Third, the control class had more game time than the treatment class. This was due to students in the control class not performing peer assessment, as thus, leaving more time to play. Last, because the lesson content was identifical for the two classes, the control class benefitted from the data

collected in the treatment class. The teacher used the assessment data to plan lessons for the two classes. Future research may need to control the practice and game time of both groups. At the onset of the study, the researcher planned to have 10 hours of lessons, but eventually had to settle for eight hours. Also, the researcher could not dictate the class size, as the class allocation was done by the school. There were four badminton courts which could at maximum allow 16 students to play at any one time assuming four students play in each court. The control class had 19 students allowing ample space for all students to practice at the same time by making using of the spaces in between the courts. In the treatment class, the teacher had to creatively come up with ways to engage all students with a class size of 30. That being said, the response rate of the students in treatment class remained significantly higher than that of the students in the control class.

Conclusion

Formative assessment is an essential component of classroom work, and its implementation can raise standards of achievement (Black & Wiliam, 1998). The research questions for this case study were: a) How does the incorporation of AfL tools affect students' acquisition of skills in a badminton unit? b) How does the incorporation of AfL tools in a badminton unit impact students' engagement (i.e., response rate)? The treatment class used an analytical rubric and other AfL strategies to receive feedback on their performance while the control class did not. Although students in the control class had a slightly higher success rate than the treatment class, the difference was not statistically significant. Both classes had nine students showing improvement in their level of performance according to the rubric, despite the control class having more game time. This showed that the importance of incoporating Assessment for Learning tasks into lessons to improve students' performance, and not just

providing practice time alone. This case study also showed that AfL has a significant impact on students' engagement, but no significant impact on their skill acquisition. When AfL strategies were included in the lessons, students were more engaged in their learning, as compared to students in classes where no assessment strategies were adopted.

So, is Assessment for Learning worth the effort? Teachers often say that they do not have time to assess. This study showed that assessment is possible during instruction. Students had time to learn, and they received feedback on their performance. Teachers can at the same time use the assessment data to document the learning process and inform future instructions. In the current study, the skill acquisition and engagement was only measured during game play, future studies could look at the students' responses throughout the lesson (including drills and practice time). This would be able to provide a better picture of students skills in open and closed setting. Although in this study, there was not any significant difference in the skills acquisition when AfL was incorporated into lessons, AfL is still worth the effort as there was higher student engagement, and teachers could use the data to make informed decisions on lesson progression. As this study was only carried out for eight lessons, the researcher strongly believes that with more time, there would be significant skill level improvement as students were more engaged during the lessons and that every single student received quality feedback on his/her performance. As schools usually face large class sizes and space contraints, incorporating AfL tools could circumvent this issue as half of the students would be peer-assessing while the other half performing. It also allowed peer-assessers to be meaningfully and cognitively engaged when peer assessing, instead of having students wait for their turn to play.

Assessment for Learning when incorporated in lessons can increase engagement and motovation. Future research could look at the length of time on each task, to get the optimal

engagement from students. Although the results of this Case Study could not be generalized, it contributed to the limited research of the impact of assessment for learning in physical education.

References

Assessment Reform Group. (2002). Assessment for learning. Retrieved from http://www.assessmentforlearning.edu.au/default.asp

Black, P., & Wiliam. D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan 80(2),* 139–44.

Brink, M., & Bartz., D., (2017). Effective use of formative assessment by high school teachers. *Practical Assessment Research & Evaluation, 22(8),* 1-10.

Chappuis, S., Commodore, C., & Stiggins, R. (2017) *Balanced Assessment Systems: Leadership, quality and the role of classroom assessment*. Thousand Oaks, CA: Sage Publications

Cheah, Y. (1998). The examination culture and its impact on literacy innovations: the case of Singapore. *Language and Education, 12(3),* 192-209.

Chng, L., & Lund, J., (2018). The impact of formative assessment on the success and response rate in a sixth-grade physical education badminton unit. *International Journal of Kinesiology in Higher Education,* DOI: 10.1080/24711616.2018.1509680

Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: teachers' perspective and experiences. *Physical Education and Sport Pedagogy, 18(2),* 219-233.

Crouch, D., Ward, P., & Patrick, C. (1997) Effects of peer-mediated accountability on task accomplishment during volleyball drills in elementary physical education. *Journal of Teaching in Physical Education, 17,* 26-39.

French, K., Rink, J., Rikard, L., Mays, A., Lynn, S., & Werner, P. (1991). The effect of practice progressions on learning two volleyball skills. *Journal of Teaching in Physical Education, 10,* 261-274.

Gopinathan, S. (2001). Globolisation, the state and education policy in Singapore in Tan, J., Gopinathan, S., Ho, W. (Eds.) *Challenges facing the Singapore education system today*. Singapore: Prentice Hall.

Hastie, P. (1998). Effect of instructional context on teacher and student behaviors in physical education. *Journal of Classroom Interaction, 33(2),* 23-31.

Hay, P. (2006). Assessment for learning in physical education. In *The handbook of physical education,* ed. D. Kirk, D. Macdonald and M. O'Sullivan, 312-25. London: Sage.

Hellison, D. (1985). *Goals and strategies for teaching physical education*. Champaign, IL: Human Kinetics.

Holt, J., Kinchin, G., & Clark, G. (2012) Effects of peer-assessed feedback, goal setting and group contingency on performance and learning by 10-12 year old academy soccer players. *Physical Education and Sport Pedagogy, 17(3),* 231–250.

Houston-Wilson, C., Dunn, J., Van der Mars, H., & McCubbin J. (1997) The effect of peer tutors on motor performance in integrated physical education classes. *Adapted Physical Activity Quarterly, 14(4),* 298-313.

Johnson, M., & Ward, P. (2001). Effects of class-wide peer tutoring on correct performance of striking skills in 3rd grade physical education. *Journal of Teaching in Physical Education, 20,* 247-263.

Leirhaug, P. & Annerstedt, C. (2016) Assessing with new eyes? Assessment for learning in norwegian physical education. *Physical Education and Sport Pedagogy, 21(6),* 616-631.

Lim-Ratnan, C. (2013). Tensions in defining quality pre-school education: the Singapore context. *Educational Review, 65(4),* 416-431.

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

Lund, J., & Veal, M. (2013). Assessment-driven instruction in physical education. Champaign, IL: Human Kinetics.

MacPhail, A., & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': Students' and teachers' experiences of assessment for learning in post-primary physical education. Assessment in Education: Principles, Policy & Practice, 17(1), 23-29. Doi:10.1080/09695940903565412.

Ministry of Education (MOE) Singapore (2013). Holistic assessment to support new PE syllabus. Retrieved from https://www.moe.gov.sg/news/press-releases/holistic-assessment-to-support-new-pe-syllabus

Ministry of Education (MOE) Singapore, (2016). The Physical Education Syllabus 2014. Retrieved from https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/physical-sports-education/files/physical_education_syllabus_2014.pdf

Ministry of Education (MOE) Singapore (2018). Learn for life – preparing our students to excel beyond exam results. Retrieved from https://www.moe.gov.sg/news/press-releases/-learn-for-life---preparing-our-students-to-excel-beyond-exam-results

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

Oterio-Saborido & Gonzalez-Jurado (2015). Design and validation of a tool for formative assessment of invasion games. *Journal of Physical Education and Sport, 15(2),* 254-263.

Popham, J. (1997). What's wrong - and what's right - with rubrics. Association for Supervision and Curriculum Development (ASCD). EBSCO Publishing.

Tan, K. (2016). Asking questions of (what) assessment (should do) for learning: the case of bite-sized assessment for learning in Singapore. *Educational Research Policy Practice*. DOI: 10.1007/s1067-016-9196-5.

Tannehill, D., Van der Mars, H., & MacPhail, A. (2013). *Building effective physical education programs*. Sudbury, MA: Jones & Bartlett.

Van der Mars, H., McNamee, J., & Timken, G. (2018). Physical education meets teacher evaluation: Supporting physical educators in formal assessment of student learning outcomes. *The Physical Educator, 75*, 581-615.

Veal, M. (1988). Pupil assessment perceptions and practices, *Journal of Teaching in Physical Education, 7(4),* 327-342.

Vroom, V. (1964). *Work and Motivation*. New York: Wiley.

Ward, P., Crouch, D., and Patrick, C. (1998a) Effects of peer-mediated accountability on opportunities to respond and correct skill performance by elementary school children in physical education. *Journal of Behavior Education, 8(1*), 103-114.

Ward, P., Smith, S., Makasci, K., & Crouch, D. (1998b). Differential effects of peer mediated accountability in elementary physical education. *Journal of Teaching in Physical Education, 17,* 470-478.

Yan, Z. & Cheng, E. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teacher and Teacher Education, 45,* 128-136

APPENDICES

Appendix A

List of Articles Selected for Review

Author (s)	Country	Participant / Setting	Method	Purpose / Theme	Major Findings
Brink & Bartz (2017)	USA	3 teachers	In-depth case study over 3 years	Teachers' perception of formative assessment	The staff development had an especially positive impact on the teachers' understanding and skill sets for individualizing instructional practices. Support for formative assessment by the administrative team members was essential to creating a cultural shift from summative to formative assessment.
Butler & Hodge (2001)	USA	24 students in a high school (9th grade).	Open-ended questionnaire (PAQ) that was given to each student	Students' perception of formative assessment	Three themes emerged: (a) type of feedback (general or specific), if any, peers provided their partner after conducting the assessment, (b) the perceived importance of providing feedback to peers, and (c) developing trust.
Chróinín & Cosgrave (2013)	Ireland	2 Primary teachers in Limerick and 3 in Dublin, taught in a range of class from 5-12 years.	Qualitative analysis of interviews	Teachers' perception of formative assessment and impact of formative assessment	The inclusion of assessment in physical education lessons provided structure and focus to the planning, teaching and learning processes and impacted positively on both teacher learning and the children's learning. The assessment strategies focused the learners, allowed for feedback related to assessment criteria and informed future planning.
Crouch et al., (1997)	USA	67 elementary students grades 4 – 6.	Single-case: Withdrawal	Impact of formative assessment	Students performed more trials and were generally more successful during the peer-mediated accountability conditions than the other 2 conditions.
Hastie et al., (2012)	USA	20 students from two 3rd grade classes	Use IOA	Validating formative assessment tool	Overall, the results indicate the students in this study were minimally successful at assessing their own skill performance during one episode of dribbling regardless of the type of recall.

Author (s)	Country	Participant / Setting	Method	Purpose / Theme	Major Findings
Haug & Fischer (2015)	Germany	320 students from 9-13 years.	Questionnaire	Students' perception of formative assessment and impact of formative assessment	When grades are removed, students show far more instrinsic motivation in their learning. Students have taken greater owership over their development and become more accountable for their own learning.
Hill & Miller (1997)	USA	Two classes of 5th grade students (N = 54)	IOA. The Pearson product-moment correlation test was used.	Validating formative assessment process	Results indicated that there were significant correlations between the scores of the student and adult recorders for all five tests.
Johnson & Ward (2001)	USA	11 third grade students	Single case: multiple baseline	Impact of formative assessment	During intervention, children performed fewer total trials and more correct trials, and a higher % of correct trials than baseline. Intervention was similarly effective for lower and higher skilled girls. Organization of lesson time unaffected. Students accurately determine each others' performance.
Kolovelonis & Goudas (2012)	Greece	48 fifth- and sixth-grade students randomly assigned to 3 experimental groups	Quantitative: ANCOVA	Impact of formative assessment	The results showed that students were moderately accurate in peer- and self-recording with a tendency to overestimate their performance. No difference among the 3 experimental groups in recording accuracy was found. Students who received more accurate feedback outperformed in the chest pass test those who received less accurate feedback.
Leirhaug & Annerstedt (2016)	Norway	23 PE teachers across 6 upper secondary schools in Norway	Mixed Method design	Implementation of formative assessment	For the majority of the students in the study, their reports of assessment practice in PE did not reflect the four key principles of AfL.
Leirhaug & MacPhail, (2015)	Norway	3 PE teachers selected from the study above	Qualitative Case Study	Implementation of formative assessment	Teachers know about AfL but the implementation was constricted. The need of embedding AfL in learning theory is one of the strongest challenges to enacting AfL in physical education.
MacPhail & Halbert (2010)	Ireland	20 schools nominated 1 teacher each	Qualitative	Teachers' perception and impact of formative assessment	Teachers: better planning, easier to manage and organise lesson, assessment wheel easy to administer, teachers

Author (s)	Country	Participant / Setting	Method	Purpose / Theme	Major Findings
					highligh motivated and energised. Students more engaged in learning and more reflective.
Mintah, 2003	USA	210 Public school PE teachers	Questionnaire	Teachers' perception and implementation of formative assessment	Authentic Assessment found to be used extensively in public schools, i.e. Teacher observation, self-observation, checklists, peer observation, event task etc. were the most popular. Portfolio and essay were the least commonly used techniques. PE teachers perceived that AA use enhanced positively the self-concept, motivation, and skill achievement of their students.
Otero-Saborido & Gonzalez-Jurado (2015).	Spain	22 students of fourth level of primary education	IOA data	Validation of formative assessment tool	Very high IOA. Tool validated.
Otero-Saborido, et al., (2015).	Spain	62 sixth grade pupils from a state school in Spain.	Quantitative	Validation of formative assessment tool	The positive results obtained from observer precision reliability, reinforce the possibility of using this tool as a method of assessment in primary education.
Penney et al., (2012)	Australia	5 teachers, 72 students in 4 schools.	Qualitative	Teachers' and students' perception of formative assessment	Students have perceived the assessment task to be authentic and meaningful for the Physical Education Studies course and have liked the way in which 'practical' and 'theoretical' aspects are combined in the task. Teachers have identified the task as aligning well with the pedagogic intent of the course and as providing a valid means of assessment of students' skills, knowledge and understandings relating to the aspects of course content that it was designed to address.
Redelius & Hay (2012)	Sweden	A total of 355 students from 28 different schools participated. They were 15 to 16 years of age	Questionnaire, focus group	Students' perception of formative assessment	The results indicated that students do think grades are important but they did not appear to recognise the official criteria as the predominant basis for achievement of grades in PEH. Significantly, the degree of student certainty in these elements was underpinned by their indication that the grading criteria were clear and that

Author (s)	Country	Participant / Setting	Method	Purpose / Theme	Major Findings
		and attending school year 9.			they were aware of the basis upon which grading judgments were made.
Richard et al., (1998)	USA	6 elementary PE teachers teaching from different schools (grades 5-8).	Mixed Method	Validating formative assessment tool	Reaction towards the integration and usefulness of TSAP were very positive and informative. However, students must be trained and reminded of possible variables. Overall, TSAP was good for tactical games.
Veal & Compagnone, (1995).	USA	151 sixth-grade students enrolled in eight classes in three different middle schools	Mixed method	Impact of formative assessment	This study produced inconclusive results regarding differences between the experimental and comparison classes. As a motivational tool, effort is a critical component on which teachers should focus if enhanced skill is the desired outcome. Study did not show that formative assessment has an effect on self-perception. Formative assessment also needs to be linked to what will eventually be used for summative assessment.
Ward et al., (1998a)	USA	24 students in fourth Grade	Single case: Withdrawal	Impact of formative assessment	Students perform more trials and were more successful in the peer-mediated accountability condition than the other 2 conditions.
Ward et al., (1998b)	USA	9 Students from fourth and fifth Grades	Single case: multiple baseline	Impact of formative assessment	Peer mediated accountability was effective in increasing opportunities to respond for both average and low skilled students, but did not change the percentage of correct performances. Peer mediated accountability was effective to promote OTR but inappropriate when students cannot perform the skill.
Yan & Cheng (2015)	Hong Kong	450 teachers from 10 primary schools	Questionnaire	Teachers' perception	The results showed that instrumental attitude, subjective norm, and self-efficacy were significant predictors of teachers' intentions to conduct formative assessment. Teachers' intentions were most strongly predicted by self-efficacy and strongly predicted by instrumental attitude

Appendix B

MOE Singapore Games Rubric for Badminton

		Stages of Progression			
		Beginning	**Developing**	**Achieved**	**Exceeding**
Indicators	**Serve**	Serve over the net with only one type of serve	Serve low / high over the net and into the service box	Serve low / high to create space at rearcourt / forecourt [LO2, LO3] in relation to where the opponent is standing	Serve low / high to the corners of the service box to create space, in relation to where the opponent is standing
	Return Shot	Hit the shuttle over the net using predominantly one type of shot	Hit the shuttle over the net using two different shots	Hit the shuttle into open space, away from the opponent, using a variety of shots [LO1, LO4, LO5, LO6, LO7] *(e.g. net shot, net lift, overhead clear, drop shot, smash)*	Place the shuttle to the corners and open space to increase scoring opportunities
	Space Coverage	Move to react to the shuttle	Move to base position after every contact with the shuttle	Move to a strategic base position* after every contact with shuttle with proper footwork, ready for the next shot [LO8] **Varies according to where the shot was placed*	Move to provide wide coverage of the court with effective footwork to prevent scoring opportunities

Appendix C.

Unit and Assessment Plans of the Badminton Unit

Class A – Experimental Group (with AfL activities)

Lesson Number	Lesson Activities	Assessment Plans
1	• Introduction – rules of game and court lines • Pre-test: Game Play (10 minutes), singles game, full court, random partner • Training students to peer assess (use tally counting form) • Show Badminton rubrics to get students to peer-assess. Inform them of the expected learning outcomes.	Pre-Unit Assessment using rubrics by 2 Independent Teachers Tally count (10-minutes game play).
2	• Teach Overhead clears • Game Play – half court, singles	Peer-assessment – number of overhead clears
3	• Overhead clears practice • Underhand lifts • Game Play – half court, singles	Peer-assessment – number of Underhand lifts
4	• Teach net shots • Game play – Half court, singles	Peer-assessment – number of net rallies
5	• Teach Footwork • Game play – practice all skills taught	Peer-assessment – Tally count (10-minutes game)
6	• Game Play (half court, singles) • Review where students place their shots and which are the shots they missed.	Peer-assessment – Heat map
7	• Game Play (half-court, singles) • Review where they place their shots and space coverage.	Peer-assessment – Heat map
8	• Game Play (half-court, singles) • Ladder Competition	Self-assessment – Rubrics Student reflection Post-unit assessment using rubrics by 2 Independent Teachers

Class B – Control class

Lesson Number	Date Time	Lesson Activities	Assessment Plans
1		• Introduction – rules of game and court lines • Game Play (10 minutes), singles game, full court, random partner • Show Badminton rubrics to explain the learning outcomes. Students to self-assess and rate where they think they are	Pre-Unit Assessment using rubrics by 2 Independent Teachers
2		• Teach Overhead clears • Game Play – half court, singles	
3		• Teach underhand lifts • Overhead clears practice • Game Play – half court, singles	
4		• Teach net shots • Practice overhead clears • Game play – half court, singles	
5		• Teach Footwork • Game play – half court, singles	
6		• Game Play (half court, singles) • Ladder Competition	
7		• Game Play (half-court, singles) • Ladder Competition	
8		• Game Play (half-court, singles) • Ladder Competition	Post-unit assessment using rubrics by 2 Independent Teachers

Badminton Unit

Formative Assessment – Return Shots Tally

Performer: _____( ) Class: _____

Observer: _____( ) Class: _____

Instructions

A successful return shot in badminton is when the shuttle goes over the net and lands within the back half of the court.

Get a partner to observe you playing a half-court game for 10 minutes. Your partner should tally the shots in with the 'Successful' or the 'Unsuccessful' column. Tally marks look like this: ~~////~~ //.

Personal Recording Form – Return shots

	Successful	**Unsuccessful**
Tally		
Total:	(a)	(b)

Calculate % of successful shots $\frac{(a)}{(a) + (b)}$ X 100% = _____ %

Tally Percentage from Lesson 1: _____ %

Compare your tally percentage today to the percentage in Lesson 1. Write down what you think about the two results and write an improvement goal for yourself.

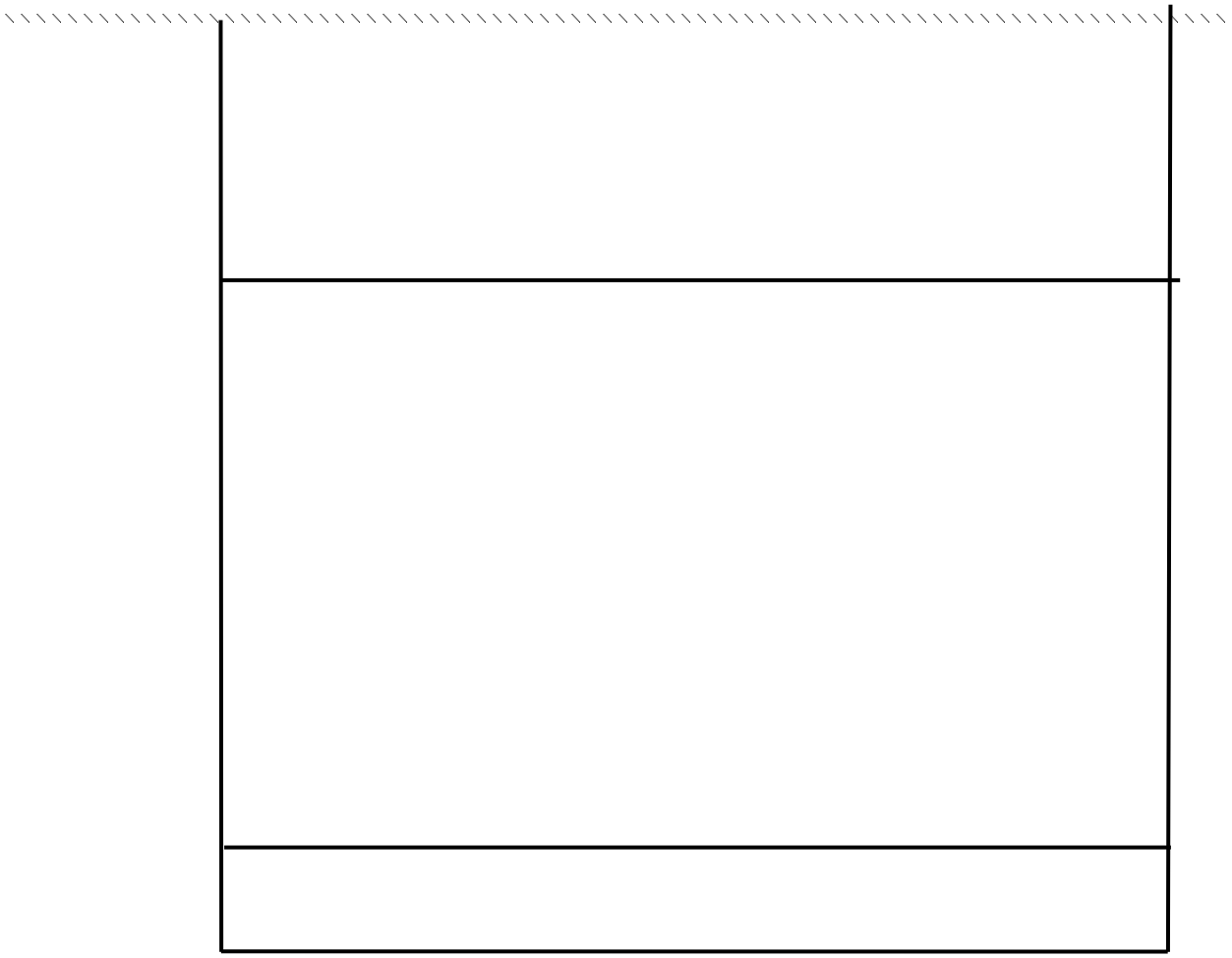
Heat map for Badminton

Player: _____ **Opponent:** _____

Observer: _____

Task:

- Watch your partner play a singles (half court) game for 10 minutes.
- Mark on the court below, a ‘S’ on where the serve of the shuttle lands.
- Mark on the court below, a ‘O’ when your partner successfully returns the shot; and an ‘X’ when your partner fails to return the shot.



Self-reflection

1. Write on the court the word ‘BASE’ on your heatmap to show where you should be standing after returning every shot.

2. Look at the areas on your own side of the court where you managed to return the shuttles successfully. Is it a wide area? Are you satisfied with you results? Why and Why not?

3. Look at the areas you **<u>did not</u>** manage to return the shuttle. What can you do to improve?

4. Look at your opponent’s heat map. State if you have hit the shuttle to:
    a. Front of the court near the net Yes / No
    b. Left side of the court Yes / No
    c. Right side of the court Yes / No
    d. Back of the court Yes / No

5. Look at your opponent’s heat map. State if you have served the shuttle to:
    a. Front of the court near the net Yes / No
    b. Left side of the court Yes / No
    c. Right side of the court Yes / No
    d. Back of the court Yes / No

6. After looking at your opponent’s heat map, what can you do to improve?

7. On a scale of 1 – 5 (1 being not confident, 5 being very confident), rate how well you can execute the following shots:

Type of Shots	**Rating**
Overhead Clears / Drive	
Underhand lifts	
Tumbling net shots	
Short Serves	
Long Serves	

Appendix D

Skills Ratings by Independent Observer

Table D1

Group: Experimental Group

		Lesson 1			Lesson 8		
Number	Skills Group	**Serve**	**Return Shot**	**Space Coverage**	**Serve**	**Return Shot**	**Space Coverage**
1	LS	1	1	1	1	1	1
2	HS	1	2	1	1	2	**2**
3	MS	1	1	1	1	**2**	1
4	HS	1	2	1	**2**	**3**	**2**
5	LS	1	1	1	1	**2**	1
6	LS	1	1	1	1	1	1
7	MS	1	2	1	**2**	**3**	**2**
8	HS	1	2	2	1	2	2
9	LS	1	1	1	1	**2**	1
10	HS	2	3	2	2	3	2
11	MS	1	2	1	1	2	1
12	HS	1	2	1	1	2	1
13	HS	1	2	2	1	2	2
14	HS	1	2	1	1	2	**2**
15	HS	1	2	1	1	2	**2**
16	MS	1	1	1	1	**2**	**2**
17	LS	1	2	1	1	2	1
18	MS	1	1	1	1	1	1
19	MS	1	1	1	1	1	1
20	LS	1	1	1	1	1	1

Table D2

Group: Control class

		Lesson 1			Lesson 8		
Students	**Skills grouping**	**Serve**	**Return Shot**	**Space Coverage**	**Serve**	**Return Shot**	**Space Coverage**
1	MS	1	1	1	1	1	1
2	MS	1	1	1	1	1	1
4	HS	1	1	1	1	**2**	1
6	LS	1	1	1	1	1	1
7	HS	1	2	1	**2**	**3**	**2**
9	MS	1	1	1	1	**2**	**2**
10	LS	1	1	1	1	1	1
11	LS	1	1	1	1	1	1
12	HS	1	2	1	1	2	1
13	MS	1	1	1	1	**2**	1
14	HS	1	2	1	1	**3**	**2**
15	HS	1	2	1	**2**	**3**	**2**
16	MS	1	1	1	1	**2**	1
17	LS	1	1	1	1	**2**	1
18	LS	1	1	1	1	1	1
19	HS	1	1	1	1	**2**	1

Appendix E. Students' Inter-Observer Agreement with Researcher's

	Researcher's	**Student's**	
Number	**Succ %**	**Succ %**	**Difference**
1	59	50	9
2	70	76	6
3	69	57	12
4	78	86	8
5	64	55	9
6	28	60	32
7	0	0	0
8	66	57	9
9	57	55	2
10	86	84	2
11	74	63	11
12	75	62	13
13	74	57	17
14	85	83	2
15	83	89	6
16	63	77	14
17	71	68	3
18	73	74	1
19	0	0	0
20	61	40	21
Total	**1236**		177
Disagreement	14%		
Agreement	86%		

Appendix F. Activity Time Comparison

Table F1

Activity Time - include drills, practice, game play		
Lesson Nos.	Treatment	Control
1	19.7	15.5
2	32	29.4
3	29.1	35
4	27.7	36.5
5	23.9	37.6
6	21.1	33
7	26.3	36.2
8	33.9	30.4
Total	**213.7**	**253.6**
Average	**26.71**	**31.70**

Table F2

Knowledge Time - Teaching		
Lesson Nos.	Treatment	Control
1	16.1	10.5
2	17.2	11.4
3	14.5	10.6
4	14.3	9
5	22.3	6.5
6	26.8	15.6
7	26.8	8.1
8	14.1	8.9
Total	**152.1**	**80.6**
Average	**19.01**	**10.08**

Table F3

Total Learning Time (Knowledge & Activity)		
Lesson Nos.	Treatment	Control
1	35.8	26
2	49.2	40.8
3	43.6	45.6
4	42	45.5
5	46.2	44.1
6	47.9	48.6
7	53.1	44.3
8	48	39.3
Total	**365.8**	**334.2**
Average	**45.73**	**41.78**

Graph Representation

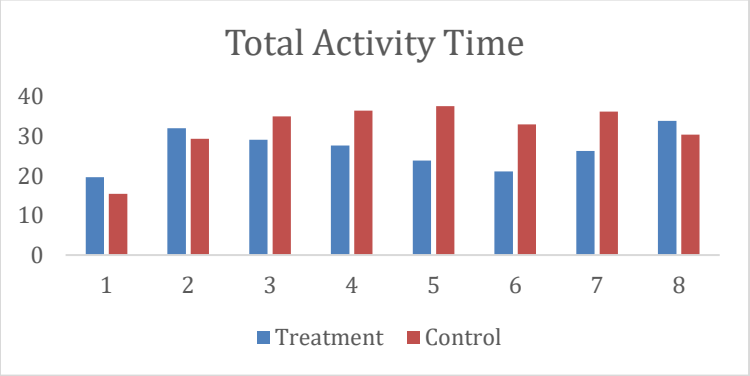


Figure F1. Total activity time between treatment and control class

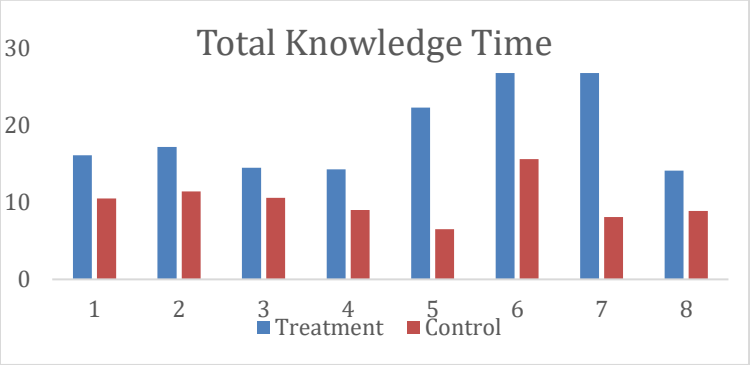


Figure F2. Total knowledge time between treatment and control class

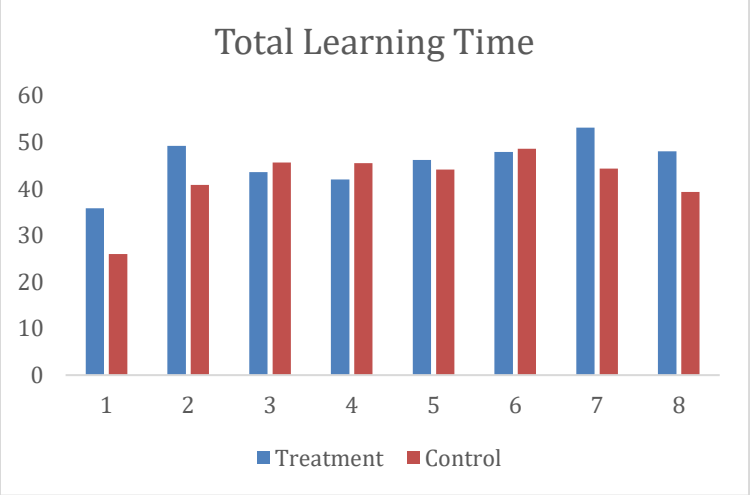


Figure F3. Total learning time between treatment and control class

Appendix G. Raw Data

Treatment Lesson 1

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	48	20	68	71	8.62
4	HS	75	19	94	80	8.4
8	HS	59	21	80	74	9.68
10	HS	46	9	55	84	8.4
12	HS	71	30	101	70	8.4
13	HS	91	37	128	71	9.68
14	HS	76	23	99	77	9.63
15	HS	78	25	103	76	8.62
Total		544		728		71.43
Average				**91**	**75%**	**8.93**
Response rate per minute	**10.19**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	37	27	64	58	8.62
5	LS	52	45	97	54	9.68
6	LS	53	54	107	50	9.68
9	LS	26	27	53	49	8.4
17	LS	38	29	67	57	8.62
20	LS	73	55	128	57	9.63
Total		279		516		54.63
Average				**86**	**54%**	**9.11**
Response rate per minute	**9.45**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	57	19	76	75	9.67
7	MS	44	20	64	69	8.62
11	MS	83	46	129	64	9.68
16	MS	62	33	95	65	9.68
18	MS	64	34	98	65	9.68
19	MS	83	26	109	76	8.62
Total		393		571		55.95
Average				**95**	**69%**	**9.33**
Response rate per minute	**10.21**					

Treatment Lesson 2

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	65	38	103	63%	9.15
4	HS	129	31	160	81%	13.8
8	HS	93	31	124	75%	10.7
10	HS	100	13	113	88%	10.7
12	HS	121	24	145	83%	13.8
13	HS	71	30	101	70%	7.7
14	HS	221	33	254	87%	19.8
15	HS	215	45	260	83%	19.8
Total		1015		1260		105.45
Average				157.5	**81%**	13.18
Response rate per minute	**11.95**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	58	35	93	62%	5.75
5	LS	29	20	49	59%	5.75
6	LS	14	46	60	23%	5.75
9	LS	67	38	105	64%	9.15
17	LS	100	32	132	76%	12.6
20	LS	24	44	68	35%	6.75
Total		292		507		46
Average				84.5	**58%**	7.63
Response rate per minute	**11.08**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	26	27	53	49%	5.75
7	MS	29	13	42	69%	5.75
11	MS	118	36	154	77%	11.10
16	MS	98	35	133	74%	12.60
18	MS	75	31	106	71%	8.80
19	MS	27	15	42	64%	6.75
Total		373		530		50.75
Average				88.33	**70%**	8.46
Response rate per minute	**10.44**					

Treatment Lesson 3

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	92	49	141	65%	12.5
4	HS	106	43	149	71%	16.75
8	HS	161	52	213	76%	16.4
10	HS	173	25	198	87%	16.3
12	HS	112	56	168	67%	16.75
13	HS	128	49	177	72%	21.2
14	HS	310	50	360	86%	22.1
15	HS	276	56	332	83%	22.1
Total		1358		1738		144.1
Average				217	**78%**	18.01
Response rate per minute	**12.06**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	0	0	0	0%	0
5	LS	62	44	106	58%	11.4
6	LS	50	173	223	22%	18.6
9	LS	93	54	147	63%	12.5
17	LS	119	41	160	74%	15.8
20	LS	0	0	0	0%	0
Total		324		636		58
Average				159	**51%**	14.58
Response rate per minute	**10.91**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	133	28	161	83%	18.6
7	MS	118	38	156	76%	15.8
11	MS	106	53	159	67%	16.75
16	MS	96	30	126	76%	15.8
18	MS	65	21	86	76%	15.8
19	MS	59	27	86	69%	8
Total		577		774		90.75
Average				129	**75%**	15.13
Response rate per minute	**8.53**					

Treatment Lesson 4

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	88	39	127	69%	16.32
4	HS	173	54	227	76%	16.28
8	HS	85	57	142	60%	16.25
10	HS	85	41	126	67%	16.25
12	HS	180	52	232	78%	16.28
13	HS	88	52	140	63%	14.33
14	HS	182	51	233	78%	17.47
15	HS	135	30	165	82%	11
Total		1016		1392		124.18
Average				174	**73%**	15.52
Response rate per minute	**11.21**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	58	30	88	66%	16.1
5	LS	126	65	191	66%	16.18
6	LS	111	52	163	68%	11.25
9	LS	58	30	88	66%	16.1
17	LS	173	69	242	71%	20.63
20	LS	74	62	136	54%	14.33
Total		600		908		95
Average				151	**66%**	15.77
Response rate per minute	**9.60**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	84	78	162	52%	16.32
7	MS	141	59	200	71%	16.18
11	MS	122	55	177	69%	16.28
16	MS	193	49	242	80%	20.63
18	MS	122	28	150	81%	11.25
19	MS	130	56	186	70%	16.28
Total		792		1117		96.94
Average				186	**71%**	16.16
Response rate per minute	**11.52**					

Treatment Lesson 5

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	64	27	91	70%	10.4
4	HS	100	28	128	78%	10.17
8	HS	65	34	99	66%	10.4
10	HS	86	14	100	86%	10.17
12	HS	103	34	137	75%	10.17
13	HS	115	40	155	74%	10.4
14	HS	150	26	176	85%	10.17
15	HS	88	18	106	83%	10.4
Total		771		992		82.28
Average				124	**78%**	10.29
Response rate per minute	**12.06**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	61	42	103	59%	10.4
5	LS	74	41	115	64%	10.4
6	LS	36	91	127	28%	10.17
9	LS	67	51	118	57%	10.17
17	LS	82	33	115	71%	10.4
20	LS	85	54	139	61%	10.4
Total		405		717		62
Average				120	**56%**	10.32
Response rate per minute	**11.58**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	75	34	109	69%	10.17
7	MS	0	0	0	0%	
11	MS	120	42	162	74%	10.4
16	MS	77	45	122	63%	10.17
18	MS	95	35	130	73%	10.4
19	MS	148	28	176	84%	10.17
Total		515		699		51.31
Average				140	**74%**	10.26
Response rate per minute	**13.62**					

Treatment Lesson 6

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	38	25	63	60%	
4	HS	89	18	107	83%	
8	HS	72	36	108	67%	
10	HS	61	17	78	78%	
12	HS	89	35	124	72%	
13	HS	109	35	144	76%	
14	HS	132	21	153	86%	
15	HS	97	27	124	78%	
Total		687		901		
Average				113	**76%**	9.60
Response rate per minute	**11.73**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	80	23	103	78%	
5	LS	55	48	103	53%	
6	LS	55	74	129	43%	
9	LS	68	39	107	64%	
17	LS	88	26	114	77%	
20	LS	0	0	0	0%	9.6
Total		346		556		
Average				93	**62%**	9.60
Response rate per minute	**9.65**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	58	35	93	62%	
7	MS	55	36	91	60%	
11	MS	101	42	143	71%	
16	MS	61	29	90	68%	
18	MS	93	28	121	77%	
19	MS	231	65	296	78%	
Total		599		834		
Average				139	**72%**	9.60
Response rate per minute	**14.48**					

Treatment Lesson 7

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	44	31	75	59%	8.77
4	HS	66	17	83	80%	8.77
8	HS	47	21	68	69%	8.77
10	HS	85	26	111	77%	8.77
12	HS	127	30	157	81%	8.77
13	HS	0	0	0	0%	0
14	HS	69	30	99	70%	8.77
15	HS	92	24	116	79%	8.77
Total		530		709		
Average				101	**75%**	8.77
Response rate per minute	**11.55**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	83	40	123	67%	8.77
5	LS	83	56	139	60%	8.77
6	LS	54	77	131	41%	8.77
9	LS	49	28	77	64%	8.77
17	LS	42	32	74	57%	8.77
20	LS	90	50	140	64%	8.77
Total		401		684		
Average				114	**59%**	8.77
Response rate per minute	**13.00**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	42	28	70	60%	8.77
7	MS	82	32	114	72%	8.77
11	MS	128	32	160	80%	8.77
16	MS	93	33	126	74%	8.77
18	MS	58	13	71	82%	8.77
19	MS	0	0	0	0%	0
Total		403		541		
Average				108	**74%**	8.77
Response rate per minute	**12.34**					

Treatment Lesson 8

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
2	HS	68	30	98	69%	16.17
4	HS	128	21	149	86%	16.17
8	HS	0	0	0	0%	16.17
10	HS	204	35	239	85%	16.17
12	HS	122	49	171	71%	16.17
13	HS	179	76	255	70%	16.17
14	HS	52	20	72	72%	16.17
15	HS	108	38	146	74%	16.17
Total		861		1130		
Average				161	**76%**	16.17
Response rate per minute	**9.98**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	LS	69	43	112	62%	16.17
5	LS	124	74	198	63%	16.17
6	LS	36	38	74	49%	16.17
9	LS	120	84	204	59%	16.17
17	LS	93	55	148	63%	16.17
20	LS	32	26	58	55%	16.17
Total		474		794		
Average				132	**60%**	16.17
Response rate per minute	**8.18**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
3	MS	88	31	119	74%	16.17
7	MS	152	45	197	77%	16.17
11	MS	202	62	264	77%	16.17
16	MS	154	37	191	81%	16.17
18	MS	31	14	45	69%	16.17
19	MS	135	75	210	64%	16.17
Total		762		1026		
Average				171	**74%**	16.17
Response rate per minute	**10.58**					

Control Lesson 1

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	198	82	280	71%	28
7	HS	246	69	315	78%	28
12	HS	169	62	231	73%	28
15	HS	195	82	277	70%	28
14	HS	0	0	0	0%	28
19	HS	252	58	310	81%	28
Total		1060		1413		
Average				283	**75%**	28
Response rate per minute	**10.09**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	36	98	134	27%	28
10	LS	102	78	180	57%	28
11	LS	47	109	156	30%	28
17	LS	52	72	124	42%	28
18	LS	94	85	179	53%	28
Total		331		773		
Average				155	**43%**	28
Response rate per minute	**5.52**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	77	51	128	60%	28
2	MS	50	33	83	60%	28
9	MS	No Data		0	0%	28
13	MS	absent		0	0%	28
16	HS	195	71	266	73%	28
Total		322		477		
Average				159	**68%**	28
Response rate per minute	**5.68**					

Control Lesson 2

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	119	49	168	71%	18.25
7	HS	197	81	278	71%	29.25
12	HS	201	77	278	72%	29.25
15	HS	187	62	249	75%	29.25
14	MS	110	28	138	80%	15.75
19	HS	188	52	240	78%	29.25
Total		1002		1351		151
Average				225	**74%**	25.17
Response rate per minute	**8.95**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	82	102	184	45%	16.5
10	LS	109	73	182	60%	18.25
11	LS	64	88	152	42%	15.75
17	LS	84	76	160	53%	17.27
18	LS	67	66	133	50%	17.27
Total		406		811		85
Average				162	**50%**	17.0
Response rate per minute	**9.54**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	108	37	145	74%	16.1
2	MS	81	37	118	69%	16.5
9	MS	76	42	118	64%	16.1
13	MS	absent		0	0%	0
16	HS	162	72	234	69%	23.27
Total		427		615		72
Average				154	**69%**	17.99
Response rate per minute	**8.55**					

Control Lesson 3

Number	Grouping	Succ	Unsucc	Total	%Succ	Game Time
4	HS	162	35	197	82%	23.15
7	HS	97	48	145	67%	18.65
12	HS	96	42	138	70%	18.65
14	HS	168	44	212	79%	22.4
15	HS	130	52	182	71%	18.65
19	HS	112	28	140	80%	18.65
Total		765		1014		120.15
Average				169	**75%**	**20.025**
Response rate per minute	**8.44**					

Number	Grouping	Succ	Unsucc	Total	%Succ	Game Time
6	LS	83	100	183	45%	17.45
10	LS	189	120	309	61%	28.5
11	LS	69	76	145	48%	23.15
17	LS	105	88	193	54%	24.3
18	LS	200	76	276	72%	24.3
Total		646		1106		117.7
Average				221.2	**58%**	**23.54**
Response rate per minute	**9.40**					

Number	Grouping	Succ	Unsucc	Total	%Succ	Game Time
1	MS	234	80	314	75%	28.5
2	MS	180	78	258	70%	30.23
9	MS	226	100	326	69%	30.23
13	MS	148	79	227	65%	28.17
16	MS	136	64	200	68%	28.17
Total		924		1325		145.3
Average				265	**70%**	**29.06**
Response rate per minute	**9.12**					

Control Lesson 4

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	189	64	253	75%	28.92
7	HS	0	0	0	0%	0
12	HS	204	86	290	70%	30.82
14	HS	285	72	357	80%	28
15	HS	305	89	394	77%	34.25
19	HS	247	62	309	80%	30.82
Total		1230		1603		152.81
Average				321	**77%**	30.56
Response rate per minute	**10.49**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	153	114	267	57%	30.6
10	LS	153	130	283	54%	31.67
11	LS	105	58	163	64%	28.92
17	LS	225	126	351	64%	31.67
18	LS	275	105	380	72%	33.17
Total		911		1444		156
Average				289	**63%**	**31.2**
Response rate per minute	**9.25**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	206	88	294	70%	28.92
2	MS	98	34	132	74%	28
9	MS	276	108	384	72%	30.6
13	MS	252	101	353	71%	34.25
16	MS	220	107	327	67%	33.17
Total		1052		1490		155
Average				298	**71%**	30.99
Response rate per minute	**9.62**					

Control Lesson 5

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	151	48	199	76%	38.5
7	HS	134	68	202	66%	38.5
12	HS	0	0	0	0%	0
14	HS	226	51	277	82%	38.5
15	HS	214	76	290	74%	38.5
19	HS	176	66	242	73%	38.5
Total		901		1210		192.5
Average				242	**74%**	38.50
Response rate per minute	**6.29**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	164	105	269	61%	38.5
10	LS	181	149	330	55%	38.5
11	LS	0	0	0	0%	0
17	LS	190	129	319	60%	38.5
18	LS	293	157	450	65%	38.5
Total		828		1368		154
Average				342	**61%**	38.5
Response rate per minute	**8.88**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	326	94	420	78%	38.5
2	MS	162	79	241	67%	38.5
9	MS	192	109	301	64%	38.5
13	MS	304	97	401	76%	38.5
16	MS	212	105	317	67%	38.5
Total		1196		1680		193
Average				336	**71%**	38.50
Response rate per minute	**8.73**					

Control Lesson 6

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	121	39	160	76%	23.33
7	HS	74	29	103	72%	23.33
12	HS	74	33	107	69%	23.33
14	HS	102	24	126	81%	23.33
15	HS	117	34	151	77%	23.33
19	HS	115	38	153	75%	23.33
Total		603		800		139.98
Average				133	**75%**	23.33
Response rate per minute	**5.72**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	104	56	160	65%	23.33
10	LS	85	47	132	64%	23.33
11	LS	65	45	110	59%	23.33
17	LS	75	64	139	54%	23.33
18	LS	70	40	110	64%	23.33
Total		399		651		116.65
Average				130	**61%**	23.33
Response rate per minute	**5.58**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	101	38	139	73%	23.33
2	MS	108	61	169	64%	23.33
9	MS	74	36	110	67%	23.33
13	MS	59	34	93	63%	23.33
16	MS	163	44	207	79%	23.33
Total		505		718		116.65
Average				144	**70%**	23.33
Response rate per minute	**6.16**					

Control Lesson 7

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	143	50	193	74%	34.5
7	HS	172	75	247	70%	34.5
12	HS	156	55	211	74%	34.5
14	HS	154	56	210	73%	34.5
15	HS	203	57	260	78%	34.5
19	HS	163	65	228	71%	34.5
Total		991		1349		207
Average				225	**73%**	34.50
Response rate per minute	**6.52**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	25	12	37	68%	34.5
10	LS	169	75	244	69%	34.5
11	LS	93	90	183	51%	34.5
17	LS	113	102	215	53%	34.5
18	LS	125	63	188	66%	34.5
Total		525		867		173
Average				173	**61%**	34.5
Response rate per minute	**5.03**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	165	65	230	72%	34.5
2	MS	162	63	225	72%	34.5
9	MS	142	86	228	62%	34.5
13	MS	144	51	195	74%	34.5
16	MS	147	45	192	77%	34.5
Total		760		1070		173
Average				214	**71%**	34.50
Response rate per minute	**6.20**					

Control Lesson 8

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
4	HS	128	34	162	79%	29.83
7	HS	117	53	170	69%	29.83
12	HS	101	15	116	87%	29.83
14	HS	157	47	204	77%	29.83
15	HS	142	45	187	76%	29.83
19	HS	200	48	248	81%	29.83
Total		845		1087		178.98
Average				181	**78%**	29.83
Response rate per minute	**6.07**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
6	LS	49	58	107	46%	29.83
10	LS	75	79	154	49%	29.83
11	LS	85	61	146	58%	29.83
17	LS	88	68	156	56%	29.83
18	LS	116	60	176	66%	29.83
Total		413		739		149
Average				148	**56%**	29.8
Response rate per minute	**4.95**					

Number	Grouping	Succ	Unsucc	Total	Succ %	Game Time
1	MS	114	57	171	67%	29.83
2	MS	97	88	185	52%	29.83
9	MS	72	39	111	65%	29.83
13	MS	111	35	146	76%	29.83
16	MS	159	38	197	81%	29.83
Total		553		810		149
Average				162	**68%**	29.83
Response rate per minute	**5.43**					

Appendix H. Inter-Observer Agreement

IOA Calculation

	Disagreement	Agreement	Total
Experiment Lesson 3	446	2773	
Experiment Lesson 6	223	2068	
Control Lesson 3	341	3104	
Control Lesson 6	124	2295	
Total	1134	10240	11374
Average 90%			

Experiment Lesson 3

	Independent Observer's			Researcher's			IOA		
Bib Nos.	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**
1	0	0	0	0	0	0	0	0	
2	77	43	120	92	49	141	15	6	
3	119	44	163	133	28	161	14	16	
4	101	32	133	106	43	149	5	11	
5	52	45	97	62	44	106	10	1	
6	122	85	207	50	173	223	72	88	
7	103	43	146	118	38	156	15	5	
8	154	46	200	161	52	213	7	6	
9	77	51	128	93	54	147	16	3	
10	171	16	187	173	25	198	2	9	
11	96	55	151	106	53	159	10	2	
12	108	54	162	112	56	168	4	2	
13	172	76	248	172	76	248	0	0	
14	293	38	331	310	50	360	17	12	
15	266	48	314	276	56	332	10	8	
16	84	41	125	96	30	126	12	11	
17	108	48	156	119	41	160	11	7	
18	56	28	84	65	21	86	9	7	
19	73	36	109	59	27	86	14	9	
20	0	0	0	0	0	0	0	0	
				2303	916	3219	243	203	446

Experiment Lesson 6

	Independent Observer's			Researcher's			IOA		
Bib Nos.	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**
1	77	43	120	80	23	103	3	20	
2	33	29	62	38	25	63	5	4	
3	56	28	84	58	35	93	2	7	
4	85	16	101	89	18	107	4	2	
5	51	47	98	55	48	103	4	1	
6	88	37	125	55	74	129	33	37	
7	53	37	90	55	36	91	2	1	
8	61	35	96	72	36	108	11	1	
9	60	36	96	68	39	107	8	3	
10	59	16	75	61	17	78	2	1	
11	96	36	132	101	42	143	5	6	
12	83	35	118	89	35	124	6	0	
13	103	38	141	109	35	144	6	3	
14	128	21	149	132	21	153	4	0	
15	91	28	119	97	27	124	6	1	
16	62	29	91	61	29	90	1	0	
17	83	29	112	88	26	114	5	3	
18	89	23	112	93	28	121	4	5	
19	216	63	279	231	65	296	15	2	
20	0	0	0	0	0	0	0	0	
				1632	659	2291	126	97	223

Control Lesson 3

	Independent Observer's			Researcher's			IOA		
Bib Nos.	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**
1	216	79	295	234	80	314	18	1	
2	165	66	231	180	78	258	15	12	
4	137	50	187	162	35	197	25	15	
6	93	76	169	83	100	183	10	24	
7	85	45	130	97	48	145	12	3	
9	214	108	322	226	100	326	12	8	
10	179	104	283	189	120	309	10	16	
11	89	48	137	69	76	145	20	28	
12	88	45	133	96	42	138	8	3	
13	144	75	219	148	79	227	4	4	
14	161	41	202	168	44	212	7	3	
15	101	53	154	130	52	182	29	1	
16	129	56	185	136	64	200	7	8	
17	101	80	181	105	88	193	4	8	
18	191	73	264	200	76	276	9	3	
19	99	29	128	112	28	140	13	1	
						3445	203	138	341

Control Lesson 6

	Independent Observer's			Researcher's			IOA		
Bib Nos.	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**	**Succ**	**Unsucc**	**Total**
1	94	42	136	101	38	139	7	4	
2	99	60	159	108	61	169	9	1	
4	111	39	150	121	39	160	10	0	
6	98	52	150	104	56	160	6	4	
7	100	74	174	100	74	174	0	0	
9	64	35	99	74	36	110	10	1	
10	76	50	126	85	47	132	9	3	
11	60	35	95	65	45	110	5	10	
12	110	78	188	110	78	188	0	0	
13	50	34	84	59	34	93	9	0	
14	98	22	120	102	24	126	4	2	
15	142	49	191	142	49	191	0	0	
16	154	45	199	163	44	207	9	1	
17	64	62	126	75	64	139	11	2	
18	71	34	105	70	40	110	1	6	
19	152	59	211	152	59	211	0	0	
						2419	90	34	124