Exploring the associations of students' intrinsic and extrinsic motivation towards high-stake tests in Physical education.

- a correlational study using Self-Determination Theory

Leonardo Hansson & Simon Riesler
Exploring the associations of students' intrinsic and extrinsic motivation towards high-stake tests in Physical education.

A correlational study using Self-Determination Theory

Leonardo Hansson & Simon Riesler

Abstract

This study explored the associations of students’ intrinsic and extrinsic motivation towards high-stake tests in Physical education. The study collected data with questionnaires from 81 Swedish compulsory school students. The questionnaires in this study were based on the Academic Motivation Scale. Furthermore, this study was based on Self-Determination Theory. In congruence with previous research, this study found that females received a significantly higher grade than male students. The results also suggest that there was no significant difference between the different test-groups actual performance. The results of the study concluded that there are no correlations between motivation [towards high-stake tests] and the actual performance of students. Moreover, students were more motivated by external regulation (receiving higher grades) than intrinsic motivation to know (to learn a skill for life). Students perceived higher test anxiety if they were to perform in front of their classmates. However, the test-groups did not perform statistically differently which then leads to the question on the necessities of using these forms of high-stake test since it does not increase performance but is perceived to be more stressful for students. Further research is necessary to draw any generalizable conclusions if there is any correlation between motivation, high-stake tests, and perceived test anxiety.

Keywords

Academic Motivation Scale, Motivation, Self-Determination Theory, High-Stake Tests, Physical Education, Intrinsic Motivation (IM), Extrinsic Motivation (EM)
# Contents

1. Introduction .................................................................................. 1

2. Background .................................................................................. 2
   2.1 Theories on motivation .............................................................. 3
   2.2 The Academic motivation scale ................................................. 3
   2.3 Physical education ..................................................................... 4
   2.4 Research on test-taking ............................................................. 4

3. Method ......................................................................................... 5
   3.1 Data ......................................................................................... 5
      3.1.1 Participants ........................................................................ 6
      3.1.2 Test-groups ....................................................................... 6
      3.1.3 Questionnaire ..................................................................... 7
      3.1.4 Validity ............................................................................. 8
   3.2 Ethical considerations ............................................................... 8
   3.4 Analysis .................................................................................... 8

4. Results ........................................................................................... 9
   4.1 Correlation between different types of motivation and performance ....... 10
   4.2 Differences between test groups in performance (narrow time frame vs longer time frame) .................................................. 10
   4.3 Correlation between perceived test anxiety and performance .................. 10
   4.4 Genders differences in performance, motivation and perceived test anxiety ........................................................................... 10
   4.5 Summary of the results ................................................................ 12

5. Discussion ..................................................................................... 12
   5.1 Different types of motivation and correlation with performance ............. 13
   5.2 Differences between test groups (narrow time frame vs longer time frame) ................................................................. 13
   5.3 Perceived test anxiety correlation with performance ............................ 13
   5.4 Gender differences in performance, motivation, and perceived test anxiety ........................................................................... 14
   5.3 Implications for teaching ................................................................ 15
   5.5 Limitations ............................................................................... 16

6. Conclusion .................................................................................... 16

References ....................................................................................... 18

Appendix A ....................................................................................... 20
List of abbreviations:
SDT - Self-Determination Theory
IM - Intrinsic Motivation
EM - Extrinsic Motivation
AM - Amotivation
AP - Actual Performance
PPP - Perceived Performance Prior
PPA - Perceived Performance After
PA - Perceived Anxiety
IFCM - Infront of Classmates
IFC - Infront of Camera
ACL - At Chosen Location
NF - Narrow Time Frame
LF - Longer Time Frame
1. Introduction

“In live as if you were to die tomorrow. Learn as if you were to live forever” - Mahatma Gandhi

There is an inherent challenge in the current educational system in Sweden as the curriculum strive for students to achieve the “lifelong desire to learn” while simultaneously directing students towards academic achievement as grades in Sweden according to Klapp (2015) directly affect a students’ opportunities when applying to higher education (Skolverket, 2018, p. 5).

Prior research suggests that “High-stake tests are intended to motivate students to perform to higher standards (Ryan, Ryan, Arbuthnot & Samuels, 2007; Nordgren, Odenstad & Samuelsson, 2017).” However, Fitzgerald (2015) found in her systematic review that high-stake testing has many negative consequences for students’ performances such as anxiety and sleep deprivation for and before the days of the tests.” In contrast to these findings, a study by Tus (2020) found no correlation between stress, motivation, and academic performance.

In school, teachers have limited time to teach students what Skolverket demands in the curriculum each year. Skolverket is the government body responsible for the Swedish school system with the aim that all who attend a Swedish school receive equal schooling with high quality in a great environment (Skolverket, 2021). A problem with the curriculum is time management, as each knowledge requirement is rarely on schedule. Teachers have to determine what parts of the curriculum in a school year are only tested once. Often, this is the case with dance examinations and the traditional way in Sweden is to show choreography in front of the class. Since there is time for only one of these exams each school year the pressure on students is immense. On top of that, students are asked to perform a choreography presented live in a high-stake setting, meaning that most of the time students have one chance to perform. The knowledge requirement on dancing states the following: “In dance, and movement and training programs to music, pupils adapt to some extent their movements to beat, rhythm and context.” (From Skolverket, 2018, p. 52). In the knowledge requirement about dance in lgr11, it is not stated how students are supposed to be examined. Even so, it is not uncommon for students to perform in front of their class which may induce anxiety and reduce performance which is not a part of the educational program. Moreover, the grades received in both compulsory school and upper secondary school have an impact on the future socio-economic status of the student after graduation and also affect the probability of the student applying for University (Klapp, 2015). This in turn affects many more aspects of the students' life. This could be compared to elite athletes where only results matter and the pressure to perform is enormous.

Self-Determination Theory (SDT) differentiates different forms of motivation (Deci & Ryan, 2000; Haerens et al., 2019). Motivation is distinguished into autonomous forms or more volitional forms, controlled or more pressured forms, and amotivation which is a lack of motivation. STD has shown that students who feel satisfaction linked to their basic psychological needs of autonomy, competence, and relatedness are more prone to
be autonomously motivated and less prone to exhibit controlled motivation or amotivation (Deci & Ryan, 2000).

There are few studies in physical education that have examined motivation and perceived performance in compulsory school students regarding high-stake tests. The studies that have examined motivation and perceived performance in Physical Education have not used the same variables to test their thesis and have not examined students’ motivation when taking high-stake tests. Research (Haerens et al. 2019) has shown that students’ situational motivation is affected by their knowledge of the assessment criteria for the upcoming test (see Haerens et al. 2019). However, Haerens and colleagues’ (2019) studies did not examine if students’ motivation would have differed in a high-stake test with a narrow time constraint in contrast to a test with length constraints. In contrast, “[... ] large-scale educational data results suggest that lower stake tests are associated with lower test-taking motivation (Eklöf & Knekta, 2017).” Meanwhile, a study by Hosch (2012) examined time on tests, student motivation, and performance, and found that more time spent on tests had a significant impact on test performance. These findings validate further examination on different forms of high stake-test taking. Although SDT is generally used in qualitative studies, many studies have utilized the theory in quantitative studies as well.

Graetz and Karim (2019) found in their report that female students generally graduated with higher grades, than their male counterparts, at the compulsory and high school level. Other studies (Ryan et al., 2007, Guay, Morin, Litalien, Valois, & Vallerand, 2015) that have researched motivation and or high-stake tests also included gender similarities and differences regarding motivation and performance. Ryan and colleagues (2007) found that female students had lower self-efficacy beliefs than male students despite performing equally or better. Guay and colleagues (2015, p. 72) did not examine self-efficacy and instead used the SDT framework and found that “women tend to be more intrinsically motivated by accomplishment, more regulated by identification and introjection, but less amotivated and regulated by external sources of control”. Therefore, gender was included in this study to explore whether the study is in congruence with previous research.

This study aims to broaden the knowledge of high-stakes tests association with students’ performance, motivation, and test anxiety. These findings could be useful for physical educators in Sweden when designing high-stake tests, leading to a less stressful testing environment.

2. Background

This background section will be divided into four sections: theories on motivation, Academic Motivation Scale, physical education, and research on test-taking. The first section will discuss theories on motivation. The second will discuss the research on the Academic Motivation Scale. The third section will discuss research on physical education. The last section will discuss research on test-taking.
2.1 Theories on motivation

Cook & Artino (2016, p. 997) defines motivation “as the process whereby goal-directed activities are initiated and sustained.” There are many different theories surrounding motivation. Therefore, it is difficult to determine which motivational theory has the most validity and reliability (Cook & Artino, 2016). Self-determination theory has a president of being used in educational contexts (Guay et al., 2015; Harenes et al., 2019; Litalien et al., 2017; Vallerand, 1992; Vasconcellos et al., 2020) measuring student motivation, whereas social cognitive theory (and, the key concept within it, self-efficacy) is not as popularised despite their somewhat overlapping intended areas of use.

Self-Determination Theory developed by Deci and Ryan (1985) refers to individuals’ motivation in terms of self-determination and free will. The theory covers three types of motivation: intrinsic motivation (IM), extrinsic motivation (EM), and amotivation (AM). Additionally, it is emphasized that the understanding of human motivation necessitates the consideration of innate physiological needs for competence, autonomy, and relatedness (Deci & Ryan, 2000).

To understand these two types of motivation, intrinsic and extrinsic, it is important to define these concepts and when they are applicable. IM is explained as students who experience autonomous forms of motivation do so when they “engage in behaviors for reasons that are perceived as self-endorsed and volitional.” (Litalien et al., 2017, p. 67). Contrary to students who experience controlled forms of motivation (extrinsic) who “engage in behaviors for reasons that are perceived as resulting from internal or external pressures, reflecting a lower sense of volition” (Litalien et al., 2017, p. 67). Amotivation can be described as a lack of motivation and absence of reason or lack of willingness to engage in specific behaviors (Litalien et al., 2017, p. 68).

2.2 The Academic motivation scale

Research examining self-determination in education used the Academic motivation scale (AMS) (Vallerand, 1992; Guay et al., 2015). The AMS has been used to measure academic motivation and has been reported to be substantial in its predictive validity in the field (Guay et al., 2015). Moreover, this is not a main part of the Self-Determination Theory. However, in the AMS, intrinsic motivation is differentiated and since our measurements are based on the AMS this is the way we measure motivation. The first posits that intrinsic motivation can be seen as a multidimensional concept (Carbonneau & Vallerand, 2012). Furthermore, the tripartite Model of Intrinsic Motivation (TMIM) identifies three types of intrinsic motivation, to know, to accomplish, and to experience stimulation (Guay et al, 2015). IM to know refers to the pleasure of learning, exploring and, understanding something new. IM to accomplish is the pleasure of trying to compete against yourself and surpassing yourself. Moreover, it is the satisfaction to create something new. Lastly, IM to experience stimulation refers to finding any sort of enjoyment in a task (i.e playing for the sake of playing) (Guay et al., 2015).

Studies researching learning outcomes using SDT in schools reveal that autonomous motivation is correlated with students’ learning and outcomes of performed tasks and students’ attitudes toward school (Litalien et al., 2017). Different types of intrinsic
motivation and extrinsic motivation as identified in SDT have been strongly correlated with autonomous student motivation (Vasconcellos et al., 2020).

2.3 Physical education

Students’ performance in Physical education in compulsory school (P.E) is mostly affected by physical literacy and motivation (Ennis, 2015). Mandigo and colleagues (2009, p. 28) define physically literate students as individuals who move with competence in a variety of activities that develop the entire body (e.g. gymnastics). Furthermore, these people also develop consistent motivation and showcase the ability to understand the concept of a movement. Where the individual can understand, explain, use and analyze these movements. Moreover, the person can also apply these movements when demonstrating and do so consistently, confidently, creatively in a variety of settings across multiple physical activities where they are deemed suitable.

Although the performance in P.E is not solely based on these two variables, they are highly important to succeed in P.E. (Ennis, 2015; Hattie, 2014; Haerens, 2019). Intrinsic motivation (IM) helps people perform better in areas increasing their knowledge and in many students’ cases their grades (Bandura, 1997). This is achieved by tailoring lessons for the individuals of the class so that students are less likely to compare themselves to their peers and focus on their progression (Bandura, 1997). Furthermore, Bandura argues that a “Self-comparison of improvement in a personalized classroom structure raises perceived capability.“ (Bandura, 1997, p. 12). In Hattie’s (2014) systematic review of learning, students’ perceived capabilities showed a strong correlation with academic performance.

In P.E students with higher autonomous motivation tend to be more physically active and engaged during classes (Haerens et al., 2019). Research has also found that amotivation is related to several negative outcomes in P.E.

2.4 Research on test-taking

Grades in Sweden are highly important for students' futures as they are used in applying for further education (i.e. upper secondary school and university) (Klapp, 2015). This inevitably means that all graded tests fall under the category of high-stake tests. High-stake tests are examinations that are required for students to for example pass and finish school, get admitted to a university, and apply for certain jobs (Kumandas & Kutlu, 2015). High-stake tests are intended to motivate students to perform to higher standards (Ryan et al. 2007; Nordgren et al. 2017). On the contrary, Fitzgerald (2015) found in her systematic review that high-stake testing has many negative consequences for students’ performances such as anxiety and sleep deprivation for and before the days of the tests. Furthermore, the review also revealed that students, especially minorities, in the U.S had a higher rate of failure and dropout chance was made at the beginning of the 1950s and the cumulative evidence indicates that perceived test anxiety and performance are negatively correlated in standardized tests such as intelligence tests and achievement tests (Schillinger, Mosbacher, Brunner, Vogel & Grabner, 2021).
A study by Hosch (2012) examined time on tests, student motivation, and performance and found that the time students spent on exams had a significant influence on performance, where students who spent more time on exams performed better than those who spent less.

Low-stakes assessment has, according to Schüttpelz-Brauns and colleagues (2020), in recent years, suffered from a highly varied test-taking effort in students. Similarly, a study in Sweden using large-scale educational data found results suggesting that lower stakes tests are associated with lower test-taking motivation (Eklöf & Knekta, 2017). However, the authors found evidence that there is a disparity between motivational levels between students and thus, motivational strategies could affect students differently.

When it comes to genders differences in student motivation, findings suggest that there either is no particular differences between the genders, that male students have reported higher self-efficacy beliefs or, that female students have a lower self-efficacy than males but they have equal or higher performance than male students (Ryan et al., 2007).

In Sweden, male students generally perform better than females at the SweSAT (i.e. a standardized test used for admission to Swedish universities) (Graetz & Karim, 2019). Interestingly, male students do not perform better in other standardized tests except Mathematics (Skolverket, 2019). This would explain why male students perform better at the SweSAT as 50% of the grade is calculated from the score received in Mathematics. However, there is no SweSAT equivalent test currently in use in P.E, therefore, it is difficult to assess if these findings also occur in P.E.

3. Method

This quantitative study aims to explore high-stake test association with students' intrinsic and extrinsic motivation and perceived performance in physical education. The study utilized questionnaires to survey compulsory school students’ motivation towards high stake test exams. This study aims to answer the following research questions:

- If different types of motivation are correlated with performance?
- If there are any differences in performance between test groups (narrow time in front of the class, narrow time frame in front of a camera, and longer time frame at a chosen location)?
- If perceived test anxiety is correlated with performance?
- If there are any gender differences in performance, motivation and perceived test anxiety?

3.1 Data

The data used in this study was collected through an online questionnaire (see appendix A for the different items) distributed in a compulsory school in Huddinge municipality
in Stockholm. We chose a convenience sample due to time constraints of this study and the difficulties in finding participants in an ongoing global pandemic. The questionnaire only utilized close-ended questions as it provides a high standardization which also improves its validity.

3.1.1 Participants
The target population consisted of Swedish compulsory school students from municipalities close to Huddinge in Stockholm. The school is a private school which means that the parents of students have to apply to be admitted, but students who live in closer proximity to the school get prioritized admission (Stockholms Stad, 2021). Furthermore, in the specializations in Art and, Music and Drama the applying students must pass tests to be accepted into the school. However, for the classes in Nature- and Social sciences no admission test is necessary.

The participants were at least 15 years old at the time of the study to ensure that they could give consent to their participation in the study. The gender make-up in Huddinge municipality consists of 50.7% male and 49.4% female (Länsstyrelsen, 2018). The participants of this study belonged to a school with profiling from the 6th grade divided into four different profiles: STEM (science, technology, engineering, and mathematics), Music and Drama, Art and Global (social sciences). Each of these correlates to different subjects that have increased the number of lesson hours in each week.

The data of the study were collected in the following manner. The students at the compulsory school were asked if they wanted to participate in the study under the prerequisite that they were 15 years old. The students were then given a link to the questionnaire through their digital platform.

This study surveyed 81 students in total, 81 in the first questionnaire, 78 in the second, and 77 in the third. There were 58 unique responses in questionnaires 1, 2, and 3. There were 67 unique responses in questionnaires 1 & 3. The total number of participants exceeds the minimum requirement of 30 for quantitative studies to attain a normal distribution according to Dörnyei (2016). In total there would have been 102 participants but 5 could not participate in the study as they were not 15 years old at the time and another 16 participants did not participate in the study as they were either: not present at these lessons due to illness etc., or the not want to participate in the study and some of the participants were taking the Swedish national tests at the time of the study.

3.1.2 Test-groups
The participants groups were randomly selected in their class and divided into three test-groups using a random group generator. Group 1 performed a dance in front of their classmates [NF] (IFCM) with a narrow time frame where on the examination day they had 5 minutes to practice and perform their dance. Group 2 performed in front of a camera [NF] (IFC) and no peers also with a narrow time frame with 5 minutes to practice and perform, and group 3 danced at their chosen location (ACL) with a longer time frame [LF] of 24 hours to practice and perform. After the tests, the students were asked to answer a questionnaire.
3.1.3 Questionnaire

This study used three questionnaires, two before the high-stake tests and one after the high-stake tests. The first questionnaire contained fewer questions and contains two sections. The first section contains a description of the questionnaire and consent form where the participants were asked if they were willing to participate in the study. The second section contains several group-related questions on the participants’ gender, compulsory school specialization that is natural sciences, music and drama, arts or global profile, what examination form the individual was randomly placed. Here the participants also were asked to provide a personal id, based on their initials and the three last digits of their phone number, this allows the study to track the participants’ responses throughout the three questionnaires while keeping the responses anonymous in the data analysis. The remaining questions in this section pertained to the participant’s prior dance experience, how frequently they have danced in the last year, how secure they were in their dance ability, how secure they were to dance (in front of a group, alone or in front of a camera), perceived performance in the upcoming test, perceived test anxiety (PA) in different forms of examination (on stage, in front of a camera or at a freely chosen location).

The second questionnaire contains two sections, the first section is identical to the first section in the first questionnaire. The second section also contains the same questions on personal id, class specialization, gender, and examination form. In addition, this questionnaire contained questions about how secure the participants were in their dance ability. Whether the dance lesson made them more confident than before in the upcoming dance examination. Furthermore, students were asked in a multiple-choice question on test motivation to choose a statement which they could relate the most to on a scale of 1-7. This questionnaire is used to examine whether the lessons: were effective in preparing the participants for the exams and affected the students’ motivation and security in their dance ability.

The third questionnaire was constructed into five sections, the first section of the second questionnaire is identical to the first questionnaire. The second section contains several background questions pertaining to the participants’ dance experience, perceived performance in the upcoming test, and what compulsory school specialization they attend that is Stem, Music and Drama, Arts, or Global profile. The third, fourth, and fifth sections contain questions regarding IM, EM, and AM respectively.

This study adopted the Academic motivation scale (AMS; Guay et al., 2015), which is one of the most used scales for measuring academic motivation. The scale has also displayed substantial predictive validity in the field (Guay et al., 2015). The AMS scale uses a seven-point Likert where 1 = strongly disagree and 7 = strongly agree (Litalien et al., 2017). The scale contains 28 items which measured seven types of motivation: intrinsic motivation (IM) (to know, IM towards accomplishment, and IM to experience stimulation), extrinsic motivation (EM) (identified regulation, introjected regulation, and external regulation and amotivation (AM) (Vallerand et al., 1992; Guay et al., 2015). This study used the AMS scale with minor modifications; reducing the total items to 14 items (two per factor) to fit the current studies’ research questions on high-
stake testing. This study utilized three questionnaires because it allowed the study to measure the participants' perceived performance both before the high-stake tests and after performing the high-stake tests. These two variables were analyzed in conjunction with the participants' actual performance to examine if there are any differences in performance, perceived test anxiety, and motivation between the different variants of high-stake testing (see list of abbreviations).

Actual performance is measured as the grade the students’ received on their high-stake test. Grades are assessed by the students’ teacher under the current curriculum in P.E provided by Skolverket (2018).

Perceived performance prior is measured as the grade students predicted they would receive prior to the examination.

Perceived performance after is measured as the grade students predicted they would receive after completing the examination.

3.1.1 Validity
Piloting research instruments are essential in quantitative research and ensure better outcomes in the particular context (Dörnyei, 2016). The questionnaire in this study was written in the students’ native language to ensure that the target group would understand the questionnaire. Moreover, the study was piloted by several students and these students did not report any particular concerns in regards to understanding the contents of the questionnaire.

3.2 Ethical considerations
This study was designed in consideration of the four ethical requirements for research (see Vetenskapsrådet, 2002). These four requirements are as follows: information, consent, confidentiality, and utilization. Before the study, the participants were firstly briefed on the study’s purpose and that data was only collected through the questionnaires which fulfill the information requirement. The participants were also informed verbally and in written form that participation in the study was entirely optional. This fulfills the consent requirement. The questionnaire was anonymous by design as it does not request any information which would risk any participants’ identity. The utilization requirement is also fulfilled as the collected data was only used for its intended research purpose.

The lesson(s) and examination(s) were mandatory except for national tests requiring students to leave the lesson(s).

The questionnaire was optional meaning that the participants could decide whether to partake in the study or not.

3.4 Analysis
The study will analyze the following:
- Correlation between different types of motivation and actual performance
- Differences between test groups in performance (narrow time in front of the class, narrow time frame in front of a camera, and longer time frame at a chosen location)
- Correlation between perceived test anxiety and performance
- Gender differences in performance, motivation and perceived test anxiety

The primary data is the participant’s measured IM and EM in high-stakes test-taking, perceived performance prior (PPP), perceived performance afterward (PPA), actual performance (AP), and perceived test anxiety (PA). Correlation analysis (spearman) was used to determine if actual performance and motivation were correlated. In this study, we used an ANOVA test to examine if there were any significant differences in actual performance between the three forms of high-stake test-taking. Furthermore, the study examined gender differences in actual performance, motivation, perceived test anxiety and dance experience. The results were gathered through the data collected from the students’ answers from the questionnaires issued and analyzed using various tests in SPSS. Dispersion measures included in the study were: mean and standard deviation. The significance level of this study has been set to p<.005.

4. Results

The result section was divided into five sections: correlation between different types of motivation and performance, differences between test groups in performance (narrow time frame vs longer time frame), correlation between perceived test anxiety, genders differences in performance, motivation and perceived test anxiety and a summary of the results.

The participants of the study (81) belonged to a compulsory school with profiling: Global (21), Stem (18), Art (21), Music and Drama (21). The participants had an average of 1.3 years of dance experience. The average female student had 2.0 years of dance experience and the male counterpart had 0.4 years of dance experience. Dance experience and actual performance (i.e. the received grade) were positively correlated (r=.-27, p<.015). The results of an ANOVA test showed significant gender differences in dance experience (p<.05) where females had more dance experience.

As seen in Table 1 there were 81 participants in total with 45 female and 36 males in the first questionnaire. In the third questionnaire, there were 77 participants in total with 44 female and 34 male participants.

Table 1. Displays the gender division.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Number of participants: Questionnaire 1</th>
<th>Number of participants: Questionnaire 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>45</td>
<td>44</td>
</tr>
<tr>
<td>Male</td>
<td>36</td>
<td>34</td>
</tr>
</tbody>
</table>
4.1 Correlation between different types of motivation and performance

As seen in table 2 EM had the highest mean of 4.37 when examining all participants and this was also found in both genders; male 4.37 and female 4.54. Amotivation was the lowest form of motivation in all participants and in both genders. Intrinsic motivation was rated in between Extrinsic motivation and AM. Intrinsic- and Extrinsic motivation was positively correlated (r=.-64, p<.001). In addition, IM and AM were negatively correlated (r=.-42, p<.001). However, EM is not correlated with AM.

These results show that the target group is predominantly motivated by EM especially external regulation.

The result of an ANOVA showed that there were no significant gender differences in any type of motivation (IM, EM, and AM). There were no correlations between motivation [towards high-stake tests] and the actual performance of students.

4.2 Differences between test groups in performance (narrow time frame vs longer time frame)

An ANOVA was used to determine whether the forms of high-stake tests were significant with performance. The ANOVA displayed no significant difference between high-stake tests and any type of performance.

4.3 Correlation between perceived test anxiety and performance

ANOVA was used to compare gender’s perceived test anxiety towards different forms of high-stake tests. In the analysis, we found that females had significantly higher perceived test anxiety than males in the ACL high-stake test (p<.012). There was no significant gender difference in perceived test anxiety on the IFC (p<.057) and no significant gender difference in perceived anxiety on the IFCM (p<.07).

4.4 Genders differences in performance, motivation and perceived test anxiety

Table 2 shows participants perceived performance prior (PPP) to the exams, perceived performance afterward (PPA), and their actual performance (AP). Females have a mean actual performance of 5.09 and males have a mean of 3.92. An ANOVA test found this difference in actual performance to be statistically significant at the p<.001.

Table 2. Displays the Mean Performance and Standard Deviation of each type of performance.

<table>
<thead>
<tr>
<th>Variables:</th>
<th>All Participants: Mean</th>
<th>All Participants: Standard</th>
<th>Male: Mean n=34</th>
<th>Male: Standard Deviation</th>
<th>Female: Mean n=44</th>
<th>Female: Standard Deviation</th>
</tr>
</thead>
</table>
Table 3 displays the performance of all participants, male and female. The table also displays how the different test-groups (high-stake setting) performed.

<table>
<thead>
<tr>
<th>Test-groups:</th>
<th>All Participants: Mean</th>
<th>All Participants: Standard Deviation</th>
<th>Male: Mean</th>
<th>Male: Standard Deviation</th>
<th>Female: Mean</th>
<th>Female: Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Test-groups</td>
<td>4.57</td>
<td>1.54</td>
<td>3.92</td>
<td>1.500</td>
<td>4.09</td>
<td>1.38</td>
</tr>
<tr>
<td>IFCM</td>
<td>4.11</td>
<td>1.64</td>
<td>3.82</td>
<td>1.54</td>
<td>3.29</td>
<td>1.72</td>
</tr>
<tr>
<td>IFC</td>
<td>4.83</td>
<td>1.34</td>
<td>4.00</td>
<td>1.41</td>
<td>5.41</td>
<td>0.94</td>
</tr>
<tr>
<td>ACL</td>
<td>4.79</td>
<td>1.59</td>
<td>4.00</td>
<td>1.66</td>
<td>5.82</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Note. The grades were changed into a numerical scale and graded on a scale of 1-6 where 1 = F and 6 = A.
4.5 Summary of the results

In summary, the ANOVA test found significant differences between gender’s performance where female students performed better than males. However, there were no significant gender differences in motivation. Lastly, female students experienced significantly higher perceived test anxiety when asked about performing at a chosen location. No significant perceived test anxiety was found between genders when asked about perceived test anxiety when performing in front of a camera and in front of the class. The result of the univariate analysis found that the interactional effects in gender x examination test-groups on actual performance were non-significant.

5. Discussion

The purpose of this study was to explore the associations of students' intrinsic and extrinsic motivation towards high-stake tests in Physical education. Moreover, we investigated if different types of motivation were correlated with performance. If there were any differences in performance between test groups (narrow time in front of the class, narrow time frame in front of a camera, and longer time frame at a chosen location). We also examined if perceived test anxiety was correlated with performance. Lastly, we examined if there were any gender differences in performance, motivation, and perceived test anxiety. In the following headings, we have discussed the results of our findings and provided a context in which they are applicable in a school setting.
5.1 Different types of motivation and correlation with performance

This study found that male students' expectancy and actual performance were not significant nor was it correlated to their form of motivation. On the other hand, female students' actual performance and perception were correlated. However, Eklöf and Knekta (2017) claim that student engagement is correlated to motivation which in contrast this study could not be found. Moreover, Eklöf and Knekta (2017) found that motivation had an impact on student grades. On the other hand, in our study, the finding suggests that this might not be the case. Students' grades were not significantly impacted by motivation or test-group. Although, female students' actual performance and perception were correlated. Eklöf and Knekta (2017) claim that student engagement is correlated to motivation which in contrast our study could not be found.

5.2 Differences between test groups (narrow time frame vs longer time frame)

There was no significant difference in actual performance between any of the high-stake test groups. The study also utilized an interaction analysis that displayed no significant differences found between the genders AP in any of the high-stake tests. The lack of significance is likely due to a lack of statistical power in the study. In order to draw any conclusions, a larger sample size would have been necessary.

In Hosch’s (2012) study, he found that time spent on studying was correlated to the actual performance of students. Furthermore, Hosch suggests that low motivated students do not perform at their highest possible level. In contrast to Hosch’s study, this study did not find any correlation between different forms of examination and performance even in the ACL group that had more test time.

5.3 Perceived test anxiety correlation with performance

The findings suggested that females have significantly higher perceived test anxiety than males in ACL and in IFC. However, in IFCM it is not. These results seem to reflect previous findings where high-stake tests were correlated to perceived test anxiety (Fitzgerald, 2015).

Previous research suggests (Ryan et al., 2007) the correlation between tests and perceived test anxiety is dependent on context. In this study, we found that perceived anxiety was not correlated with students’ actual performance. However, students experienced elevated perceived test anxiety when asked to perform in front of their class. Ryan and colleagues (2007) claim that in another context students might have been affected by the examination setting, thus impacting their actual performance. Moreover, Ryan and colleagues (2007) suggest that male students have a stronger sense of self-efficacy. In congruence with previous research, this study found that female students believe their grades to be lower than the actual performance of the female students. Conversely, male students could not perceive their level in dance thus their self-efficacy was not correlated to their actual performance as the students could not comprehend the different levels of difficulty for this assignment.
Ryan and colleagues (2007) suggest that in high-stake tests distracting thoughts could hinder the performance of students. In contrast, the tests in this study were different because students only had a minute to present their dance thus not giving them enough time to be distracted. In comparison to Ryan and colleagues (2007) where students are presented with more than one problem [to solve] and less time per problem. Students in this study had one problem to solve which could explain why there was no correlation between perceived test anxiety and performance. This study’s findings suggest that students were not hindered by test perceived anxiety which could be explained by sufficient preparation for the students to perform at their highest level.

5.4 Gender differences in performance, motivation, and perceived test anxiety

The study’s result indicated that female students performed significantly better (i.e. received a higher grade) than their male counterparts. However, dance experience was positively correlated with actual performance. In addition, an ANOVA found females to have significantly more dance experience than males.

In terms of differences between the genders, female students have a significantly higher performance than male students. Despite performing better, female students underestimate their performance (i.e. PPP) meaning that their perceived performance is markedly lower than their actual performance. This discrepancy seems to reflect Ryan and colleagues’ (2017) previous research where female students have lower self-efficacy beliefs than male students despite performing equally or even better. Whilst the metrics of this study AP and PPP are not directly equated to self-efficacy these findings could warrant further research to determine if these results reoccur in other contexts.

The result of this study displays a significant difference between the gender’s grades. This seems to align with a recent report where female students received significantly higher grades in compulsory school (Graetzb & Karim, 2019).

Ryan and colleagues (2007) suggest that stereotypes hinder performance (i.e. socio-economic status and gender). There is a stereotype in Sweden that male students often perform worse than female students in dance. This stereotype is reassured in this study, female students had a higher mean than their male counterparts. Although, more studies have to be conducted to see whether female students are better at dance and why that, if so, is the case.

Interestingly, the gender difference in the SweSAT (i.e. a standardized test used for admission to Swedish universities) could be the result of a skewed population from the male side. In Graetzb and Karim’s (2019) report the authors state that generally, only males with an above-average cognitive ability than both genders take the SweSAT whereas the female group has a wider cognitive spread resulting in males outperforming females due to the male group having higher cognitive abilities. The difference in result could be partially explained by the selection bias of the males, meaning it is not necessarily that all males perform better in the SweSAT, rather males who take the SweSAT have on average a higher cognitive ability than females who take the test. The results could differ if the widespread cognitive abilities of male candidates for
SweSAT’s were greater, that is if more males took the test the differences between genders could be smaller or even shift to advantage for females. Graetz and Karim (2019) also state that if grades from high school and the SweSAT were weighted equally then male admissions to university would rise and female admission would decrease. However, due to female students now getting into their program without the SweSAT, there is no guarantee if SweSats were weighted equally that admissions would drop. As the female students with higher cognitive abilities would likely also apply to the SweSAT meaning that the level normal distribution of cognitive abilities would increase. The high-stake tests used in this study reflect previous research as the female students performed better than male students. Since the exception is mathematics (see Skolverket, 2019) it would not have mattered whether dance was measured in a standardized test as female students perform better in all subjects within the curriculum and this is no exception. It is difficult to assess if the difference in grade between the two genders was influenced by previous dance experience, time spent practicing, or motivation. In order to draw a general conclusion, more research has to be made.

5.3 Implications for teaching

Skolverket’s (2018) purpose is for students to attain “a lifelong desire to learn” but the results of this study suggest that students are highly motivated by extrinsic factors such as achieving a high final grade rather than learning for the sake of learning.

A regular examination form in school is verbal examinations in front of the class, however, there is no evidence in the curriculum that supports this form of examination. It opens the debate whether schools are responsible for students acquiring the skill of public speaking at the risk of their performance (i.e. lower grades). Fitzgerald (2015) concludes that high-stakes tests harm students’ learning and elevate student stress levels. In this study there was no evidence of harming student learning on the other hand there was evidence that students felt a higher level of stress when the expectancy of pupils was to perform in front of their classmates.

No correlation between actual performance and perceived test anxiety was found between test-groups, although, we found that students who performed in front of a camera had a higher mean than the other two test-groups IFCM and ACL. However, students perceived a higher level of perceived test anxiety when asked to perform in front of their classmates compared to the other two test-groups (IFC and ACL). Therefore, it is difficult to suggest that performing dance examinations in front of classmates is a beneficial test-taking method, at least according to this study’s findings. It does not reflect increased performance but may result in increased perceived test anxiety for students.

There is no evidence supporting the examination form of IFCM in LGR 11 for P.E (Skolverket, 2018) but rather a tradition of examining dances this way. Furthermore, high-stake examinations IFCM in different subjects.

This study suggests that students’ motivation is significantly weighted towards extrinsic motivation. This could mean that students are more focused on receiving a high grade rather than focusing on learning. The reason for this may be the structure of the school
system where students are dispossessed of their learning opportunities if students do not perform in each task. Thus, it is far more important for students to receive a higher grade than an education where they learn skills applicable outside of school. Skolverket's admiration of students’ desire for lifelong learning is directly impacted by the challenge students face within the school system. There is a direct conflict in the way students are graded and how teachers are supposed to encourage lifelong learning.

5.5 Limitations

There is a possible limitation with not allowing all participants to participate in the three different types of tests. However, if all the participants were to perform each test they would inevitably perform better as they would have more time for practice and possibly become more comfortable with the test. Also, the participants in the [LF] ACL group have more time to practice and record their high-stake test which may affect the actual performance of those participants.

There is a possible limitation of prior knowledge of students’ dance experiences. This could result in skewed statistics which may affect the generalizability of the results. Furthermore, since the study was only done in one school the population of subjects was too small to examine potential differences between the profiling.

Since we only examined one school in a suburban area in Stockholm no generalizations can be made on a national or grander scale. Furthermore, there has not been a plethora of research examining high-stake tests in physical education nor is it common for these types of studies to examine how different time frames might affect students' performance and motivation. Thus, the study’s result could prove difficult to use to draw any firm subject transcending conclusions as we examine less commonly researched variables. Furthermore, due to illness or other circumstances, many of the students did not answer the questionnaires therefore some of the statistical power in the study was lost decreasing the generalizability of the results. If this study was done on a larger scale with more participants with the same measurements the results may have shown significant values in the analysis regarding if gender and test-group affected the student’s grade.

There are various theories regarding motivation applied in a school setting, it is not certain that a generalized outcome can be determined from this study. Due to the time limitation, we cannot include all or many different theories surrounding the subject of motivation in the classroom.

6. Conclusion

The results of this study were that there were no significant differences between the different test-groups in actual performance. Furthermore, we also found that male students are unable to perceive their actual performance accurately; however, in this study, we found a positive correlation between female students' perceived grades and actual performance. In congruence with previous research, this study found that females
received a significantly higher grade than male students. However, female students also had significantly more dance experience than males. This result is noteworthy since dance experience was positively correlated with actual performance (i.e. grade). In terms of high-stake tests we did not find any significant differences between any of these (IFCM, IFC & ACL) in any type of performance (AP, PPP & PPA).

Perceived test anxiety was not correlated to students’ grades, meaning regardless of high-stake test form, students would statistically receive the same grade. However, their perceived anxiety would be significantly lower. This finding was found across both genders. Furthermore, perceived test anxiety was not correlated with student performance.

The study’s findings suggest that students’ perceived test anxiety was higher in the IFCM test form than IFC and ACL. The students’ performance was not correlated with perceived test anxiety, meaning that performing in front of classmates induces higher test anxiety but does not increase performance. Admittedly, perceived test anxiety does not correlate with student [actual] performance, it is difficult to justify the IFCM test form as there is no significant increase in performance over IFC or ACL.

The results of the study concluded that there are no correlations between motivation [towards high-stake tests] and the actual performance of students. Both genders were most motivated by external regulation (EM), that is receiving a high grade. This means that the dance segment for students was not about learning a skill for life, rather students only cared about scoring high.

In conclusion, students’ perceived test anxiety was higher when asked to perform in front of their classmates. Although this difference was not significant when compared to the perceived anxiety when asked to perform in front of a camera or at a chosen location the actual performance was not significantly impacted either. As the test-groups did not perform statistically differently it is difficult to favor examination forms that students could perceive to be more stressful.
References


Appendix A

Questionnaire - Enkät

Vi har utgått från AMS

- Biologiskt kön
- Skolklass
- Tidigare erfarenhet av dans
- Påverkar examinationsformen din motivation i detta moment?

Reformulated AMS to examine high-stake tests

IM - to know
- För välbehaget jag upplever när jag svarat på en provfråga som jag fann intressant.
- För att plugga inför ett prov möjliggör det för mig att fortsätta lära mig om många saker som jag tycker är intressant.

IM - toward accomplishment
- För behaget jag upplever när jag sitter vid provtillfället och känner mig väl förberedd.
- För tillfredsställelsen jag känner när jag håller på att klara av ett svår uppgift på ett prov
- (För behaget jag känner när jag presterar bättre på ett prov än vad jag trodde.)

IM - to experience stimulation.
- För de intensiva känslorna jag känner när jag är i en provsituation.
- För ruset jag upplever när jag studerar inför ett prov som intresserar mig (exempelvis ett SO prov om andra världskriget)

EM - identified regulation
- Jag tror att detta prov kommer hjälpa mig förbereda mig för framtida svåra/obekväma uppgifter.
- Krävande prov kommer att hjälper mig att prestera i liknande situationer i framtiden. (fråga Magnus om en bra fråga)

EM - introjected
- För att jag vill visa för mig själv att jag kan lyckas med denna uppgift.
- För att bevisa för mig själv att jag är en intelligent person.

EM - external regulation
- För att jag vill jag vill ha ett högt slutbetyg i kursen.
- Jag vill göra bra på provet för att jag vill komma in på ett bra gymnasium senare.

AM
- Jag vet inte, jag förstod inte vad jag ska gjort på provet.
- Jag förstår inte varför jag ska göra provet och ärligt talat kunde jag inte bry mig mindre