THE IMPACT OF GAMIFIED LEARNING IN HIGHER VOCATIONAL COLLEGE PHYSICAL EDUCATION CLASSES ON STUDENTS' PHYSICAL FITNESS, MENTAL HEALTH, AND LEARNING MOTIVATION: A RANDOMIZED CONTROLLED TRIAL

FANG YU

Graduate School Srinakharinwirot University

2025

ผลกระทบของการเรียนรู้ที่ใช้เกมในชั้นเรียนพลศึกษาของวิทยาลัยอาชีวศึกษาระดับสูงต่อสมรรถภาพทางกาย สุขภาพจิต และแรงจูงใจในการเรียนของนักเรียน: การทดลองแบบสุ่มควบคุม

FANG YU

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาศาสตร์การกีฬาและการออกกำลังกาย
คณะพลศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2568

THE IMPACT OF GAMIFIED LEARNING IN HIGHER VOCATIONAL COLLEGE PHYSICAL EDUCATION CLASSES ON STUDENTS' PHYSICAL FITNESS, MENTAL HEALTH, AND LEARNING MOTIVATION: A RANDOMIZED CONTROLLED TRIAL

FANG YU

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of DOCTOR OF PHILOSOPHY

(Sport and Exercise Science)

Faculty of Physical Education, Sports and Health, Srinakharinwirot University

2025

THE DISSERTATION TITLED

THE IMPACT OF GAMIFIED LEARNING IN HIGHER VOCATIONAL COLLEGE PHYSICAL EDUCATION CLASSES ON STUDENTS' PHYSICAL FITNESS, MENTAL HEALTH, AND LEARNING MOTIVATION: A RANDOMIZED CONTROLLED TRIAL

BY

FANG YU

HAS BEEN APPROVED BY THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY IN SPORT AND EXERCISE SCIENCE AT SRINAKHARINWIROT UNIVERSITY

(Assoc. Prof. Dr. Chatchai Ekpanyaskul, MD.)

Dean of Graduate School

ORAL DEFENSE COMMITTEE

.... Major-advisor
(Assoc. Prof. Dr.Sonthaya Sriramatr)

.... Chair
(Assoc. Prof. Dr.Apiluk Theanthong)

.... Committee
(Assoc. Prof. Dr.Supaporn Silalertdetkul)

.... Committee
(Asst. Prof. Dr.Phichayavee Panurushthanon)

.... Committee
(Asst. Prof. Dr.Witid Mitranun)

Title THE IMPACT OF GAMIFIED LEARNING IN HIGHER VOCATIONAL COLLEGE PHYSICAL EDUCATION CLASSES ON STUDENTS' PHYSICAL FITNESS, MENTAL HEALTH, AND LEARNING MOTIVATION: A RANDOMIZED CONTROLLED TRIAL

Author FANG YU

Degree DOCTOR OF PHILOSOPHY

Academic Year 2025

Thesis Advisor Associate Professor Dr. Sonthaya Sriramatr

China's higher vocational education students face declining physical fitness, mental health challenges, and low learning motivation. This Solomon four-group trial examined gamified learning among 200 students aged 18 to 25 from nursing, rehabilitation, and pharmacy majors. Grounded in Self-Determination Theory, the 16-week intervention featured game-based points, activities, feedback, and rankings. Physical fitness was measured using national standards, mental health via the DASS-21, and motivation through PALMS. Five hypotheses tested pretest effects, overall effectiveness, mediation mechanisms, domain-specific patterns, and performance differences. Results showed the strongest effects on motivation, moderate effects on physical fitness, and selective benefits for mental health. The study extends Self-Determination Theory to gamified contexts, provides implementation guidelines, and supports China's education modernization goals.

Keyword : gamified learning, physical education, higher vocational education, Self-Determination Theory, physical fitness, mental health, learning motivation, Solomon four-group design

ACKNOWLEDGEMENTS

With the completion of my doctoral thesis, I am filled with mixed emotions—above all, profound gratitude.

First and foremost, I would like to express my sincere thanks to my advisor, Mr. Sonthaya Sriramatr. Your rigorous academic standards, keen scholarly insight, and selfless mentorship guided me from confusion to clarity. Every discussion and every timely email response served as an invaluable compass on my academic journey. To be your student has been one of the greatest privileges of my life.

Next, I extend my deepest gratitude to my wife. Over the past three years, while managing your own work, you have shouldered the responsibility of caring for our two children—doing the laundry, cooking, soothing them to sleep, and waking up in the middle of the night—always without a word of complaint. Every time I returned home late at night, you welcomed me with a warm cup of tea and a gentle smile. Your tenderness and strength have been the greatest support enabling me to focus on my studies. You understand my dreams better than anyone, and you have worked harder than anyone. Thank you. I love you.

I also wish to thank all the friends who generously extended their help during the experiments, data analysis, and writing process. Your support transformed this thesis from a rough draft into a complete and meaningful work.

Finally, I am deeply grateful to my parents and family for their unwavering understanding and support. The completion of this thesis is but a small milestone, while the love and companionship you have given me will remain a lifelong treasure. I dedicate this work to all those I love and deeply appreciate.

FANG YU

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1
INTRODUCTION

Background

Higher vocational education (HVE), as a crucial component of China's education system (Lai & Lo, 2006), aims to cultivate technical and skilled professionals with practical operational capabilities and vocational qualities. It holds a pivotal position in the national development strategy (Zhang, 2024). In recent years, significant progress has been made in Chinese HVE, establishing a nationwide network of vocational colleges with expanding enrollment (Fan et al., 2024) According to data from the Ministry of Education of China, as of 2023, there are 3,074 higher education institutions (HEIs) nationwide, with 1,547 being higher vocational colleges (HVC), accounting for over 50%. Regarding enrollment, 10.42 million students were admitted to regular and vocational programs nationwide, with 5.55 million (53.3%) enrolled in higher vocational (associate degree) programs. These statistics highlight the significant role and rapid development of HVE in China's education system, reflecting the urgent societal demand for skilled professionals.

Physical education (PE), as a critical component of HVE, is particularly prominent (Dexqonov, 2024; Zhi-feng, 2012). The Chinese government's 2019 "Education Modernization 2035 Plan" calls for strengthening school PE to enhance students' practical skills (Central Committee of the Communist Party of China, 2019). The "Implementation Plan for the Reform of National Vocational Education," also released in 2019 (State Council of the People's Republic of China, 2019), emphasizes the importance of PE in vocational education, proposing innovative teaching models to improve students' physical fitness (PF) and vocational qualities The "Healthy China 2030 Planning Outline," a national health strategy published in 2016, underscores the significance of physical exercise for overall health, advocating for the widespread integration of health education and physical activities in schools to enhance students' physical condition (Central Committee of the Communist Party of China & State Council, 2016).

Despite the continuous promotion of national policies and the strong emphasis of educational strategies, the actual implementation of PE in HVC faces numerous challenges, and a significant gap remains between practice and overarching policy goals.

The Triple Crisis in HVE

PF Crisis and Health Deterioration

PE plays a crucial role in HVE, with its significance and impact being undeniable (Ntoumanis, 2001; Sharipovich, 2024). It promotes students' physical and mental health (MH) through systematic exercise (Hong-yin, 2011), laying a solid foundation for their future careers (Dan, 2007). Regular physical activities enhance cardiovascular function, increase muscle strength, and improve body coordination, thereby elevating overall health levels and reducing the incidence of diseases (Beets & Pitetti, 2005; Dukh et al., 2019; Prontenko et al., 2020).

However, current reality contradicts these ideals. according to relevant surveys in China, the obesity rate among vocational students increases annually, with male students having a higher obesity rate than female students. Meng (2012) conducted a study on the 2008 and 2009 cohorts of a technical college in Inner Mongolia, revealing that the development of students' weight and height was not ideal (Meng, 2012).

Evidence suggests that heavy workloads for PE teachers, low course engagement, outdated teaching methods, and a lack of clear teaching objectives fail to attract students' attention (Yi, 2017). This indirectly results in insufficient time, intensity, and frequency of physical exercise among vocational students. Research shows that only 35.1% of vocational students exercise 3-4 times per week, with a small percentage exercising for more than 60 minutes (Zhan-ping, 2006).

Due to a lack of proper physical exercise and guidance, the incidence of chronic diseases among younger populations is increasing. A study on potential risk factors for chronic diseases among 1,962 university students in Xi'an, China, revealed that over 90% of the students had varying degrees of chronic disease risk (Quan-chen, 2013).

MH Crisis and Stress-Related Disorders

The MH issues of HVC students have garnered significant attention from the education and public health sectors (Auerbach et al., 2016; Pedrelli et al., 2015). MH, according to the WHO, is "a state of good well-being in which an individual is able to recognize his or her abilities, cope with normal stresses in life, work effectively and productively, and can contribute to the community" (Galderisi et al., 2015). These issues not only impact students' academic performance but also have profound effects on their quality of life and future development. A review of the literature reveals that the primary psychological health problems faced by vocational college students are depression, anxiety, and academic stress (Pedrelli et al., 2015).

Chinese scholars conducted a study on depression among 254 female students at a HVC in Hubei Province, finding that the depression index ranged from 0.57 to 0.7, with a prevalence rate of 33.85% (Ai-jiao, 2007). Another study by Yanli Zeng et al. (2019) on nursing students at a HVC in Sichuan Province found that 28.7% of the students suffered from depression, indicating more severe psychological issues compared to regular university students (Zeng et al., 2019).

The MH issues of vocational students also manifest in anxiety. A study conducted in Anhui Province, China, found an anxiety detection rate of 17.3% among 3,809 vocational students (Shi-gan, 2014).

Research suggests that PE can improve the anxiety levels of vocational students. A study involving 30 male students who participated in a six-week physical training program, exercising three times a week for 60 minutes each session, found that exercise had a significant impact on reducing anxiety (Jadhav, 2017) .

Critically, stress-related symptoms appear more responsive to environmental interventions than clinical anxiety or depression. According to Li et al. (2005), a study utilizing the Chinese College Stress Scale found that approximately 8% of Chinese college students exhibited high levels of stress, while 74% were classified as experiencing low levels of stress, and 18% reported no stress at all (Li et al., 2005). Wu et al. (2020) found that 22.8% of Chinese college students reported symptoms of mental disorders, with uncertainty-related stress being the most significant predictor of such outcomes (Wu et al., 2020).

Critical LM Deficits

HVC students generally lack motivation, directly impacting their learning efficiency. Research indicates that vocational students often experience insufficient motivation. The lack of motivation among HVC students is primarily attributed to dissatisfaction with talent cultivation models.

HVC students often find that their teachers' methods and approaches fail to stimulate their interest in learning (Simón-Chico et al., 2023; Xing, 2011). Similar issues are evident in PE classes, the course content is often monotonous, and student motivation is lacking (Shou-d, 2008). Traditional PE methods, which primarily rely on lectures and passive learning, have faced increasing criticism from experts and skepticism from students. These methods fail to engage students effectively and stimulate their interest.

The Vicious Cycle: Insufficient physical activity, frequent MH problems, and weak LM are interwoven, creating a vicious cycle in which each factor reinforces the others. As a result, student participation in physical activities is continually reduced, and the overall health status of students is progressively deteriorated.

Although policies and traditional teaching reforms continue to advance, the multiple crises of PF, MH, and motivation remain unresolved in practice. An innovative teaching model that combines educational value with engagement is therefore urgently required.

Gamified Learning (GL) as a Solution

Self-determination theory (SDT), as one of the most influential frameworks in motivation research (Ryan & Deci, 2000; Ryan & Deci, 2020), provides a robust theoretical foundation for understanding how GL interventions operate. SDT posits that human motivation exists along a continuum from amotivation to intrinsic motivation, with the satisfaction of three basic psychological needs autonomy, competence, and relatedness serving as the essential nutrient for psychological growth, optimal functioning, and sustained behavioral engagement (Ryan & Deci, 2017). The alignment between SDT principles and gamification mechanisms makes this theoretical framework particularly valuable for designing and evaluating educational interventions in physical education contexts (Gao, 2024; Luarn et al., 2023).

The theoretical congruence between SDT and gamification is evidenced by how game elements systematically address each of the three basic psychological needs. Recent meta-analytic evidence demonstrates that gamified interventions significantly enhance students' perceptions of autonomy (standardized mean difference = 0.42) and relatedness (standardized mean difference = 0.38), with more modest effects on competence (X. Li et al., 2024). This pattern of differential effects across psychological needs provides both opportunities and challenges for intervention design, highlighting the necessity of theory-driven approaches rather than superficial application of game elements (Gao, 2024). Moreover, SDT's distinction between intrinsic and extrinsic motivation offers a crucial framework for understanding when gamification supports lasting behavioral change versus when it may inadvertently undermine intrinsic interest through overreliance on external rewards (Ryan & Deci, 2024).

In PE contexts, GL addresses the need for autonomy through multiple design features. By offering diverse learning pathways, personalized task choices, and self-paced progression systems, gamified environments enable students to exercise meaningful control over their learning processes (Atmaja & Mandyartha, 2020; Kam & Umar, 2024). Meta-analyses confirm that compared with traditional teacher-directed approaches, gamification more effectively fulfills students' autonomy needs and demonstrates superior capacity for sustaining learning motivation over time (L. Li et al., 2024). However, the functional significance of choice—a key SDT principle often overlooked in gamification research—suggests that autonomy support depends not merely on providing options but on ensuring those options align with students' interests and values (Gao, 2024). For instance, avatar customization and multiple progression paths may enhance autonomy only when students perceive these choices as meaningful rather than superficial.

The need for competence is addressed through gamification's emphasis on progressive challenge structures, immediate performance feedback, and visible achievement systems. Empirical studies demonstrate that in both anatomy and PE contexts, well-designed gamified interventions substantially improve students' perceived competence and learning engagement (Latre-Navarro et al., 2024; Sotos-Martínez et al., 2024). Game mechanics such as points, badges, and leaderboards function as competence-informational feedback when they communicate genuine skill progression rather than arbitrary rewards (Ryan & Deci, 2020). Recent research in physical activity interventions indicates that reward-feedback mechanisms (e.g., points systems and performance badges) demonstrate superior efficacy in promoting moderate-to-vigorous physical activity compared to interventions emphasizing only social elements (Wang et al., 2025). However, meta-analytic evidence reveals that competence improvements through gamification are less pronounced than gains in autonomy and relatedness, potentially due to implementation challenges such as inappropriate difficulty calibration or students' unfamiliarity with game elements undermining their sense of mastery (L. Li et al., 2024).

GL strengthens the need for relatedness through collaborative tasks, team-based competitions, and social comparison mechanisms. These features markedly enhance cooperation and social connectedness among students (Sotos-Martínez et al., 2024). Importantly, gamified group interactions serve a dual function: they not only improve relatedness but also buffer against anxiety commonly triggered by competition and assessment in traditional PE settings (Fernandez-Rio et al., 2022). The theoretical principle of mutually supportive psychological needs—another underexplored SDT concept in gamification research—suggests that satisfaction of relatedness can facilitate competence development through peer scaffolding and collaborative problem-solving (Gao, 2024). Furthermore, when social comparison mechanisms in gamified environments are designed to encourage mutual support rather than pure competition, they stimulate engagement while fostering belonging, thereby avoiding the amotivation and disengagement associated with ego-threatening competitive dynamics (L. Li et al., 2024).

The mechanistic pathway through which GL influences educational outcomes can be theoretically articulated through SDT's organismic integration framework. According to this framework, when gamified activities satisfy basic psychological needs, they facilitate the internalization of extrinsic motivation, transforming externally regulated behaviors into more autonomous forms of engagement (Ryan & Deci, 2017). This internalization process is particularly crucial in PE contexts, where long-term behavioral maintenance depends on developing intrinsic interest in physical activity rather than reliance on external contingencies. Empirical evidence supports this theoretical mechanism: gamified physical education interventions guided by SDT principles demonstrate not only immediate improvements in motivation and participation but also sustained behavioral change extending beyond the intervention period (Fernandez-Rio et al., 2022; Sotos-Martínez et al., 2024).

Within the specific context of Chinese higher vocational education, where students face the triple crisis of declining PF, mental health challenges, and motivational deficits, SDT-based GL offers a theoretically grounded solution that addresses these interconnected problems simultaneously. The theory predicts differential effects across outcome domains: strongest impacts on motivation (through direct need satisfaction), moderate effects on PF (through sustained behavioral engagement), and more selective benefits for mental health (through stress reduction rather than clinical symptom relief). These predictions align with preliminary evidence suggesting that gamification's psychological benefits, mediated by enhanced game performance and engagement, create a positive feedback loop that maintains participation and facilitates skill development over time (Luarn et al., 2023). By anchoring intervention design in SDT principles—ensuring that game elements genuinely support rather than undermine psychological needs—this study seeks to optimize the motivational architecture of gamified PE and provide empirical evidence for the theoretical mechanisms through which GL operates in vocational education settings.

Significance of the Research

This study addresses a critical gap in vocational education research by providing empirical evidence on the role of gamification in tackling the interwoven challenges faced by HVE. It contributes across theoretical, practical, and policy dimensions while maintaining strong alignment with national educational development strategies.

Theoretical Significance

First, this study extends the application of SDT to GL, offering empirical support for the theoretical mechanisms linking game elements, the satisfaction of basic psychological needs, and subsequent behavioral outcomes. It further examines how gamification influences students' motivation, engagement, and health outcomes through specific pathways.

Second, by simultaneously examining multiple outcome domains, the study reveals differentiated patterns of effect sizes, providing theoretical grounds for predicting the relative effectiveness of gamification across various educational goals. This deepens the understanding of when and how gamification exerts its impact.

Third, the study explores game performance as a mediating mechanism between psychological benefits and sustained participation, shedding light on how gamification maintains its efficacy in long-term behavioral change. This advances theoretical understanding beyond a simple "verification of effectiveness" toward a more nuanced "explanation of mechanisms."

Practical Significance

At the practical level, this study provides guidance for the integration of gamification in vocational PE, including the selection of game elements, the design of progressive structures, assessment strategies, and expectations for outcomes over time. The findings enable educators to make evidence-based decisions in curriculum design and resource allocation.

In addition, the study offers quantitative benchmarks of effect sizes, enabling administrators and teachers to set realistic expectations and evaluate program effectiveness based on empirical standards rather than subjective impressions.

Furthermore, by identifying the student characteristics and motivational profiles most responsive to gamification interventions, the study provides evidence for targeted implementation strategies, thereby optimizing the use of limited educational resources.

Policy and Societal Significance

At the policy level, this study aligns closely with several of China's national strategies, including Education Modernization 2035, the Implementation Plan for the Reform of National Vocational Education, and the Healthy China 2030 Planning Outline (Central Committee of the Communist Party of China & State Council, 2016, 2019; State Council of the People's Republic of China, 2019). By proposing innovative, evidence-based teaching models, the study supports the dual objectives of improving both educational quality and student health.

At the societal level, improvements in the physical and MH of vocational college students contribute to cultivating a healthier and more productive workforce while reducing healthcare costs associated with lifestyle-related diseases, thereby serving broader national economic goals.

Finally, this study highlights the leadership role of Chinese vocational education in integrating educational technology and advancing evidence-based pedagogical innovation, demonstrating its potential to shape international educational practice and policy.

Research Questions and Objectives

Specific Research Objectives

Objective 1: To examine the effects of GL on PF outcomes in vocational students, with particular attention to cardiovascular endurance, muscular strength, flexibility, and body composition.

Objective 2: To evaluate the impact of GL on MH indicators, focusing on stress-related symptoms, anxiety, and depression among higher vocational college students.

Objective 3: To assess the influence of GL on multiple dimensions of LM, including intrinsic motivation, enjoyment, psychological engagement, and sustained participation in PE activities.

Objective 4: To investigate game performance as a potential mediating mechanism linking MH improvements with sustained LM and engagement.

Objective 5: To establish evidence-based effect size benchmarks and implementation guidelines for the application of GL in vocational education contexts.

Hypotheses

Based on SDT, prior gamification literature, and the specific challenges of vocational education, the following hypotheses are proposed. GL is expected to produce differential effects on PF, MH, and LM among HVC students:

H1: Pretesting exerts no significant influence on PF, MH, or LM, thereby confirming the integrity of the Solomon four-group design and excluding testing effects.

H2: Compared with control groups, students receiving GL interventions exhibit significant improvements in PF, MH, and LM, with the strongest effects on motivation, followed by PF, while MH improvements remain more selective.

H3: Game performance scores significantly mediate the relationship between improvements in MH and increases in LM, with game engagement and performance positively associated with motivation.

H4: The effects of GL are domain-specific: PF improvements are concentrated in cardiovascular endurance, muscular strength, and flexibility; MH improvements are primarily reflected in stress reduction; LM enhancements are centered on intrinsic motivation.

H5: Students with higher game performance scores demonstrate superior outcomes in MH and LM, whereas PF outcomes may display different patterns due to resource allocation.

Definition of Terms

Gamified Learning

GL refers to the application of game elements and mechanics in educational settings to enhance student engagement and motivation (Deterding et al., 2011; Landers, 2014; Sailer & Homner, 2020).

Physical Education

PE is an important part of the school education system, which is not only about the training of sports skills, but also includes many aspects such as health, fitness and teamwork(Coulter & Ní Chróinín, 2013).

Physical Fitness

PF refers to an individual's ability to demonstrate vitality and efficiency in performing daily activities, as well as characteristics and abilities associated with a low risk of developing early-onset motor diseases (i.e. those associated with insufficient physical activity) (Pate, 1988).

Mental Health

According to the WHO, MH is "a state of good well-being in which an individual is able to recognize his or her abilities, cope with normal stresses in life, work effectively and productively, and can contribute to the community" (Galderisi et al., 2015).

Learning Motivation

LM is the intrinsic or extrinsic force that drives individuals to engage in learning activities. It can be categorized into two main types: intrinsic motivation and extrinsic motivation (Ryan & Deci, 2000). Intrinsic motivation arises from an individual's interest in the learning content, curiosity, or need for self-fulfillment (Halamish et al., 2019; ROMAN & FELEA, 2022). This type of motivation is often related to personal self-efficacy, interest, and achievement goals (Filgona et al., 2020; Salajegheh & Hoseinyshavoun, 2018). Extrinsic motivation, on the other hand, is driven by external factors such as rewards, social recognition, or avoidance of punishment (Nikulina &

Snezhkova, 2019). High levels of LM are typically associated with better learning outcomes and academic achievement (Sailer & Homner, 2020).

Scope of the Study

The Solomon four-group design was first proposed by American psychologist R. L. Solomon in 1949. It aims to investigate the impact of a specific factor on participants' behavior (Solomon, 1949). The fundamental principle of this experimental design is to reduce the influence of external factors, such as testing effects and historical effects, by randomly assigning participants to different groups (Braver & Braver, 1988).

The Solomon four-group experimental design not only eliminates confounding factors while exploring causal factors but also enhances the reliability and validity of the experiment (Kirisci et al., 2020).

In the field of education, Thai scholar Sriramatr, S. (2014) employs the Solomon four-group experimental design to assess the effectiveness of internet-based interventions, grounded in social cognitive theory, in promoting extracurricular physical activity among female university students in Thailand. The study demonstrates the significant impact of these interventions (Sriramatr et al., 2014). Similarly, Şanlı, E., & Ersanli, K. (2021) utilize the Solomon four-group design to evaluate the effects of logotherapy on reducing negative identity perceptions and enhancing positive identity perceptions (Şanlı & Ersanli, 2021). In clinical trials, Salim et al. (2020) employ the Solomon four-group design to investigate the impact of pain management education on nurses' knowledge and attitudes. The findings indicate a positive influence of the educational intervention on both knowledge and attitudes (Salim et al., 2020).

Therefore, this study investigates the effects of introducing GL methods in PE on students' PF, MH, and LM in Chinese HVE. The research focuses on first-year students at Nanchang Health Vocational and Technical College, class of 2024. A randomized controlled trial using a Solomon four-group design is employed, dividing participants into four groups: pre-test + intervention + post-test (E1), pre-test + no intervention + post-test (C1), no pre-test + intervention + post-test (E2), and no pre-test

+ no intervention + post-test (C2). Over a 16-week semester, the GL project is implemented in PE classes to explore its effectiveness in enhancing physical health, increasing participation in PE, improving MH, and stimulating LM.

Research Framework

Gamification theory is a strategy that applies game design elements to non-game environments, aiming to enhance user engagement and motivation. The core elements of gamification include points and rewards, leaderboards, badges, tasks and challenges, and feedback mechanisms (Deterding et al., 2011). These elements motivate users to repeat specific behaviors and increase their engagement by providing immediate feedback and a sense of accomplishment.

SDT emphasizes that individuals' motivation levels significantly increase when their needs for autonomy, competence, and relatedness are satisfied. GL, through game elements such as points, badges, and leaderboards, enhances students' autonomy (choosing their learning content and methods), competence (gaining a sense of achievement by completing tasks), and relatedness (building connections through interaction and competition with classmates) (Ryan & Deci, 2000).

As shown in Figure 1, the core intervention of this study is a carefully designed gamified PE system, incorporating a football league–style point mechanism, eight cyclic game activities (e.g., name-tag tearing, obstacle running), real-time feedback, and team-based ranking. The participants are 200 students aged 18–25 from nursing, rehabilitation, and pharmacy majors at Nanchang Health Vocational and Technical College, who take part in a 16-week intervention.

The Solomon four-group design, regarded as the *gold standard* of experimental research, is employed. Participants are randomly assigned to four groups (50 per group), including gamified intervention groups and traditional teaching control groups, while controlling for pretest effects. Three primary outcome variables are assessed: PF (using the National PF Standard), MH (measured by the DASS-21 scale), and LM (measured by the PALMS scale). Game performance scores are hypothesized to play an important mediating role in these relationships.

Five progressive hypotheses are proposed. In addition, the study examines how game performance mediates the observed improvements and explores differentiated benefit patterns among students with varying performance levels.

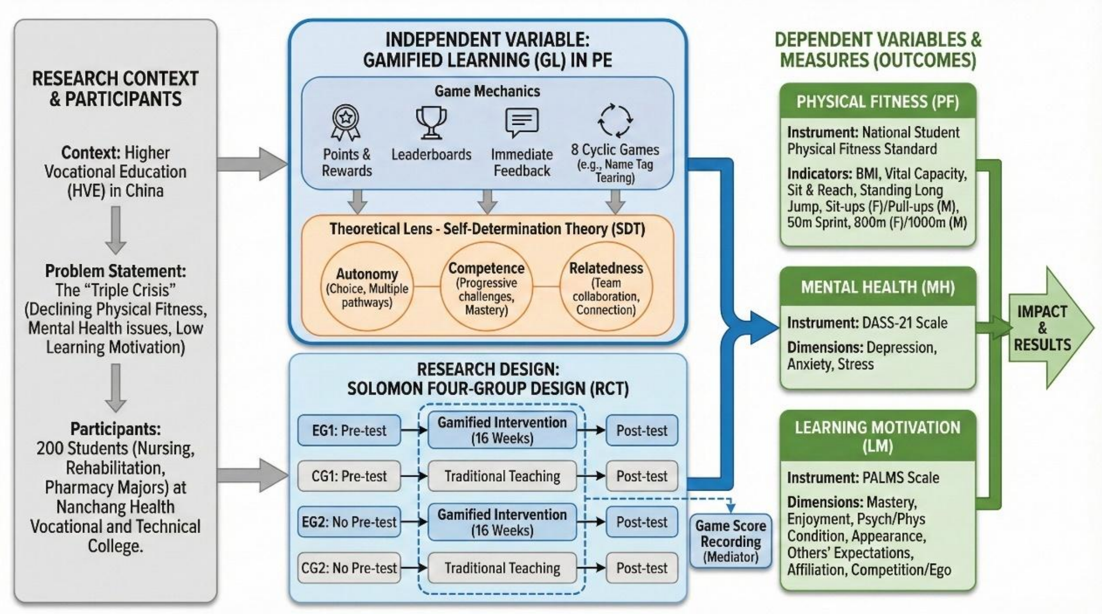


Figure 1 Basic research framework

CHAPTER 2
LITERATURE REVIEW

Theoretical Foundations for GL

SDT and Predictions of Differential Effects

SDT provides the primary theoretical framework for understanding why gamification produces differential effects across outcome domains. Its three basic psychological needs—autonomy, competence, and relatedness—predict responsiveness to gamified interventions depending on the directness of need satisfaction mechanisms (Deci & Ryan, 2012; Vansteenkiste et al., 2004).

Recent meta-analytic evidence shows that gamification effects vary substantially across outcomes: the largest effect size is observed for intrinsic motivation (d = 0.89), followed by moderate effects on behavioral engagement (d = 0.64), and smaller yet meaningful effects on physiological outcomes (d = 0.43) (Sailer & Homner, 2020). This hierarchy aligns with SDT predictions that outcomes directly tied to psychological need satisfaction show stronger responses than those requiring behavioral mediation or complex psychological processes (Kam & Umar, 2024).

Autonomy and LM: Game elements such as choice provision and multiple success pathways directly satisfy autonomy needs, producing immediate motivational benefits. Consistent evidence shows that autonomy-supportive environments yield large effect sizes for intrinsic motivation (L. Li et al., 2024; Meng & Ma, 2015).

Competence and Physical Performance: Competence needs are met through progressive challenges and achievement recognition; however, improvements in PF require sustained behavioral engagement to yield measurable outcomes. This mediating process explains why gamification studies on physical performance typically report moderate rather than large effect sizes (L. Li et al., 2024).

Relatedness and Stress Reduction: Team-based gamification elements directly address relatedness needs by fostering supportive social environments that mitigate evaluative anxiety. This mechanism primarily reduces environmental stress rather than clinical MH symptoms, thereby accounting for the selective effects on MH outcomes (L. Li et al., 2024).

Game Mechanics Theory and Engagement Processes

Game mechanics theory identifies specific elements that sustain engagement, including point systems, progressive challenges, immediate feedback, social comparison, and achievement recognition, each operating through distinct psychological pathways with characteristic temporal dynamics (Deterding et al., 2011).

Point Systems and Immediate Reinforcement: Immediate feedback mechanisms produce rapid motivational shifts within 2–4 weeks, whereas significant behavioral and physiological changes typically require 8–12 weeks (Hamari et al., 2014).

Progressive Challenge Systems: Appropriately calibrated challenges maintain optimal arousal and build self-efficacy through mastery experiences, supporting competence satisfaction and long-term participation (Kam & Umar, 2023).

Social Comparison and Competition: Team-based gamification creates controlled competitive environments that reduce evaluation anxiety common in traditional PE, while preserving the motivational benefits of social comparison (Aston, 2018; Teixeira et al., 2020).

Constructivist Learning Theory and Vocational Skill Development

Constructivist principles emphasize active knowledge construction through experiential learning, providing strong theoretical support for gamification in vocational education. Gamified environments scaffold guided practice, enabling students to construct understanding rather than passively receive information (Pascu, 2023; Riedl & Schelten, 2013).

Application to Skill Development: Evidence from medical education shows that GL environments allow students to practice clinical reasoning and decision-making in low-stress settings before applying these skills in high-pressure professional contexts (Kinio et al., 2019). Similarly, constructivist pedagogies in vocational education have been shown to integrate theory and practice in authentic tasks, thereby strengthening problem-solving skills (Van Bommel et al., 2012).

Enhancing Active Learning: Constructivist approaches emphasize learner agency and meaningful engagement with content. Gamification operationalizes these principles through interactive challenges, requiring students to solve problems actively and apply skills rather than passively absorb knowledge. Empirical studies highlight that combining gamification with constructivist strategies effectively enhances LM, collaboration, and sustained cognitive processing (Budiman et al., 2025; Zin et al., 2024).

Application of GL in HVE

Regarding the application of GL in HVE, I will elaborate on its effects on LM and MH based on a review of relevant literature.

Professional Skill Development Through Gamification

Wang Dan's research on the implementation of the Augmented Reality (AR) application "XploreRAFE+" in English classes has sparked students' curiosity (Wang et al., 2021). This study demonstrated that gamified interventions could successfully engage vocational students who typically show low motivation for traditional academic subjects, providing direct evidence for gamification's potential in vocational contexts.

Wang Fan et al. (2017) noted that Chinese HVC students are highly dependent on mobile phones and passive in classroom learning. By implementing gamified experiences in the "Distribution Center Layout and Management" course, they observed significant reform effects (Wang et al., 2017). The study's emphasis on practical skill development through game-based approaches directly parallels the current study's focus on physical skill acquisition through gamified PE.

In South Africa, a study on computer programming courses using Digital Game Based Learning found that students' academic performance surpassed that of traditional teaching methods (Yu et al., 2018). Results showed that students in gamified conditions significantly outperformed traditional instruction groups on both theoretical knowledge and practical programming skills, demonstrating gamification's effectiveness for applied learning outcomes relevant to vocational education contexts.

Jayalath & Esichaikul (2022) developed a gamified blended learning framework for technical and vocational education training, incorporating online courses with gamified content (Jayalath & Esichaikul, 2022). Their findings revealed enhanced motivation and engagement among vocational students, with particular effectiveness for students in health-related programs—directly relevant to the current study's population.

Samah, et al. (2022) used the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model in HVE training, finding that the gamified group demonstrated significantly better comprehension (Samah et al., 2022). Additionally, a study on object-oriented programming (OOP) courses implemented gamification and found that 83.48% of students met good standards, achieving teaching objectives through game incentives. The literature indicates that in HVE and training, gamification involves using information technology to design relevant programming content or frameworks, embedding games to improve student learning outcome.

MH Benefits in Vocational Settings

GL also positively impacts students' MH in HVE. At the 4th International Conference on Big Data and Education in 2021, researchers implemented GL in an English class, reducing students' embarrassment and anxiety (Zhang & Chen, 2021). An international study on a gamified software engineering education learning system evaluated students' anxiety levels and found that the GL environment significantly reduced students' learning anxiety (C.-H. Su, 2016). This pattern suggests that gamification primarily addresses situational stress related to academic performance rather than broader psychological symptoms requiring clinical intervention.

The well-known "Kahoot!" platform was used in a study with Jordanian university students, where a four-month experiment showed that anxiety symptoms in the experimental group decreased. GL not only alleviates anxiety but also benefits students with depression. A gamified meditation course recruited 72 students, and the results indicated a significant reduction in depression levels after the intervention.

Critical Analysis of MH Effects: The selective MH improvements observed across studies align with stress and coping theory, which emphasizes that environmental interventions most effectively address stressors directly related to the intervention context (academic/ PE stress) rather than broader psychological symptoms requiring clinical intervention.

Research Hotspots and Trends in Educational Gamification

Bibliometric Analysis of Gamification Research

This study utilizes CiteSpace to conduct a bibliometric analysis of literature related to GL in education from the Web of Science core database. By mapping keyword co-occurrence and conducting cluster analysis, this study aims to identify the main research hotspots and trends in GL within the educational field. Expected outcomes include clarifying the research hotspots of GL, revealing future development trends, and proposing potential future research directions.

Annual Analysis of Research Findings

The annual distribution of literature serves as a quantitative basis for delineating research stages and can generally predict future research trends in the field. This study analyzes literature on gamification in education from the period 2011 to 2024 (Figure 2).

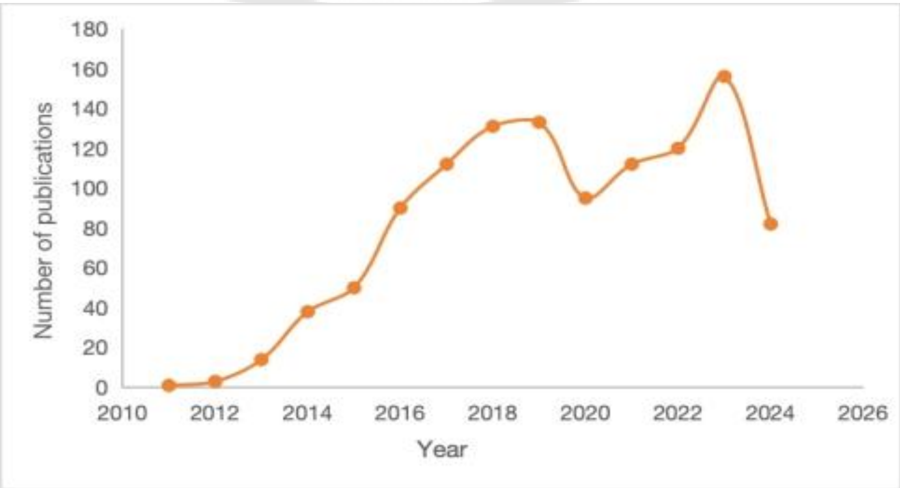


Figure 2 Annual Distribution of GL Literature in Education (2011-2024)

The number of publications related to a particular topic reflects the academic community's interest in that field. The more publications there are, the more active the research is. Based on the annual distribution of literature, the concept of GL was first introduced by Deterding et al. at the 15th International Academic MindTrek Conference in Tampere, Finland, in 2011 (Deterding et al., 2011). This marked the beginning of gamification in education, and the research in this area has grown rapidly since then.

In 2012, the number of publications increased to three. By 2013, there was a noticeable rise, with 14 publications. In 2014 and 2015, 38 and 50 publications were released, respectively, indicating a growing interest in GL in education.

From 2016 to 2019, research on GL entered a period of rapid growth. Traditional teaching methods were facing challenges such as low student engagement and insufficient motivation. GL, by introducing game elements like points, badges, and leaderboards, effectively addressed these issues and was widely adopted (Rubín, 2015). The number of publications in 2016 reached 90, increasing to 112 in 2017. In 2018 and 2019, 131 and 133 publications were released, respectively, marking a peak in research activity. This trend aligns with the global application of GL in educational practice, especially for its significant effects on enhancing student motivation and engagement (C. Su, 2016).

In 2020, the global pandemic led to a surge in the demand for online education. Although the number of publications on GL slightly decreased, it remained high, with 95 publications. The European Union's Horizon 2020 program (Commission, 2019) and UNESCO's emphasis on improving education quality through technological innovation (Rieckmann, 2017) directly influenced the publication numbers in 2021 and 2022, with 112 and 120 publications, respectively, showing steady growth.

In 2023, research on GL reached a new peak, with 156 publications. This growth is likely linked to the rapid development of online education technology, the digital transformation of education, and the advancement of AI technology (Tenório et al., 2020). As of 2024, 82 publications have already been released, and it is expected that the total number for the year will remain high.

Co-occurrence Analysis of Keywords in Publications

Keywords provide a concise summary of research topics. In literature related to a specific theme, frequently occurring keywords usually represent research hotspots in that field (Seomun et al., 2022). According to CiteSpace analysis, keywords with higher betweenness centrality indicate a higher frequency of co-occurrence with other keywords, highlighting representative terms of a research field and summarizing its key themes (Chen, 2017; Ding et al., 2001).

A co-occurrence network map of keywords was generated (see Figure 3). Excluding keywords directly related to gamification and education, keywords with a co-occurrence frequency above 10 were selected (see Table 1). The top keywords by frequency are: motivation (127), higher education (112), engagement (104), students (96), intrinsic motivation (78), performance (73), design (67), classroom (64), impact (51), and technology (38). These high-frequency keywords indicate the primary focus and trends in current research on GL in education. Keywords with a betweenness centrality greater than 0.1 include: motivation (0.10), highlighting motivation as a core theme in GL research, and technology (0.13), underscoring the crucial role of technology in this research area. These keywords act as significant mediators in the network, representing core themes in the field.

Based on these high-frequency keywords, research hotspots can be categorized into the following main areas, as shown in Table 1:

LM and Engagement: Keywords such as motivation (127), intrinsic motivation (78), engagement (104), students (96), and student engagement (29) indicate researchers' focus on enhancing students' LM and classroom engagement through gamification elements.

Educational Applications: Keywords such as higher education (112), classroom (64), and flipped classroom (26) reflect studies on the application and effectiveness of GL in higher education and flipped classroom settings.

Assessment of Learning Outcomes: Keywords like performance (73), impact (51), and achievement (24) relate to evaluating how GL affects students' academic performance and learning outcomes.

Technology and Design: Keywords such as design (67), technology (38), framework (33), self-determination theory (31), and model (27) show a focus on designing effective GL environments and the application of various technologies (e.g., augmented reality, virtual reality).

Other Related Topics: Keywords including active learning (21), perceptions (18), system (16), educational technology (16), and behavior (14) cover other important aspects of GL, such as active learning, student perceptions, and behavioral changes.

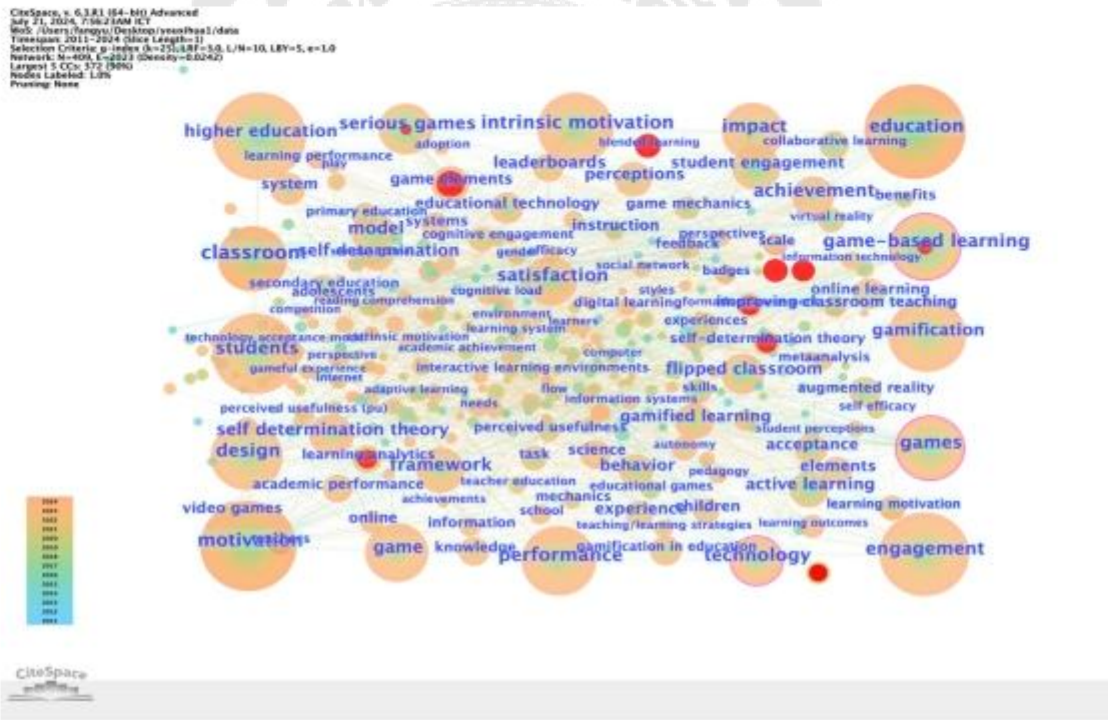
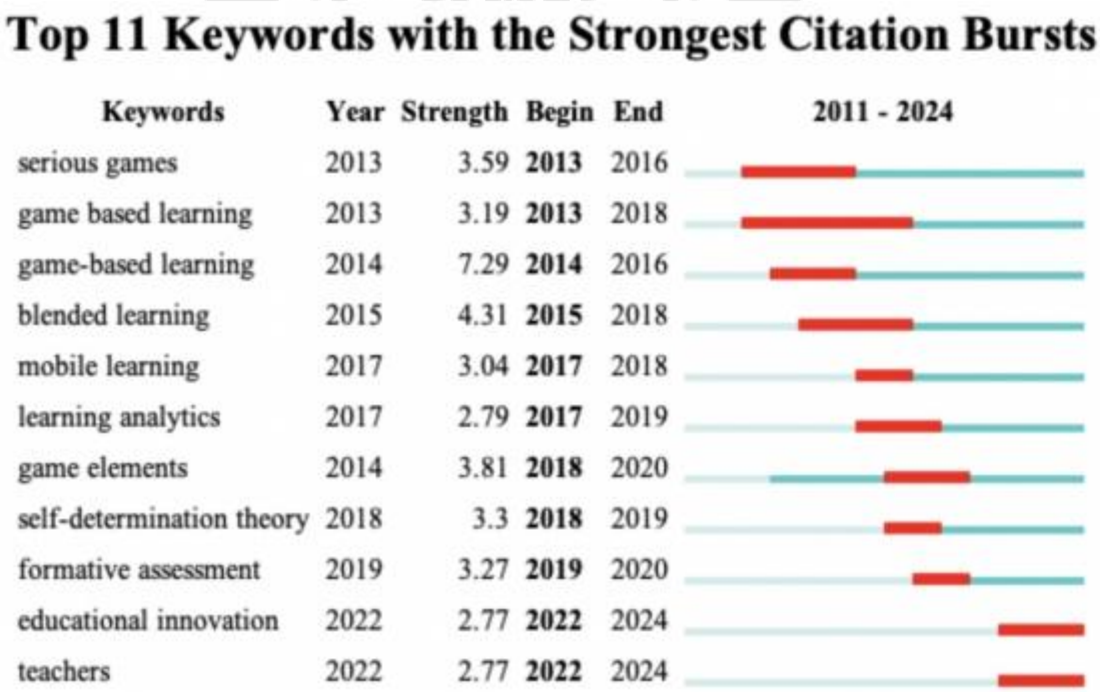


Figure 3 GL in the teaching research in the field of keywords co-occurrence knowledge map

Table 1 GL high frequency keywords statistics the research articles in the field of education

Serial number	Keyword	Frequency	Centrality	Year
1	motivation	127	0.10	2013
2	higher education	112	0.06	2015
3	engagement	104	0.04	2014
4	students	96	0.09	2016
5	intrinsic motivation	78	0.08	2016
6	performance	73	0.05	2016
7	design	67	0.07	2014
8	classroom	64	0.08	2016
9	impact	51	0.05	2018
10	technology	38	0.13	2016
11	satisfaction	36	0.06	2016
12	framework	33	0.06	2017
13	self-determination theory	31	0.03	2017
14	student engagement	29	0.03	2015
15	model	27	0.06	2016
24	online learning	15	0.02	2020
25	science	14	0.02	2019
26	leaderboards	14	0.02	2019
27	elements	13	0.02	2020
28	self determination	13	0.03	2018
29	augmented reality	13	0.02	2015
30	learning analytics	12	0.01	2017
31	knowledge	12	0.01	2018
32	collaborative learning	12	0.00	2017
33	academic performance	12	0.01	2018
34	systems	11	0.01	2022
35	experience	11	0.04	2018
36	virtual reality	11	0.03	2016
37	game mechanics	10	0.01	2014

Burst words are keywords whose frequency increases significantly over a period, representing research hotspots and trends during that time (Moreira et al., 2017). The burst word analysis map (Figure 4) shows a total of 11 burst words. "Serious games" and "game-based learning" are the earliest and longest-lasting burst words. "Game-based learning," "blended learning," and "game elements" have significant influence. The emergence of "educational innovation" and "teachers" indicates the continuous development of information technology in the educational field, particularly the advent of AI technology. It suggests that educators must keep pace with new technologies and integrate them into GL.


Top 11 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2011 - 2024
serious games	2013	3.59	**2013**	2016	
game based learning	2013	3.19	**2013**	2018	
game-based learning	2014	7.29	**2014**	2016	
blended learning	2015	4.31	**2015**	2018	
mobile learning	2017	3.04	**2017**	2018	
learning analytics	2017	2.79	**2017**	2019	
game elements	2014	3.81	**2018**	2020	
self-determination theory	2018	3.3	**2018**	2019	
formative assessment	2019	3.27	**2019**	2020	
educational innovation	2022	2.77	**2022**	2024	
teachers	2022	2.77	**2022**	2024	

Figure 4 GL in the field of education research literature mutational word analysis map

Cluster Analysis of Keywords in Publications

Keyword clustering analysis helps provide a comprehensive understanding of the hotspot structure and content within a research field (Börner et al., 2003). Using the LLR algorithm (Chen et al., 2010), a clustering analysis was conducted on the keywords in research related to GL in education, generating a cluster map (see Figure 5) and a report (see Table 2).

The Modularity (Q value) of the clustering is 0.3613, higher than 0.3, indicating a significant clustering structure. The Mean Silhouette (S value) is 0.7445, above 0.5, suggesting the clustering results are reasonable and well-formed. Table 2 shows the cluster silhouette values and cluster sizes. Cluster #0, "serious games," is the largest cluster with the most nodes, indicating that gamification has significantly enhanced classroom teaching and has been widely applied in higher education in recent years. Clusters with high silhouette values (>0.95) include "#4 teaching methods," "#5 distance learning," "#6 physical literacy," "#7 augmented reality," "#8 active learning," and "#9 game development." As shown in Figure 5, these clusters exhibit overlapping and interconnected states, indicating they are closely related and represent significant research branches in GL within the educational field.

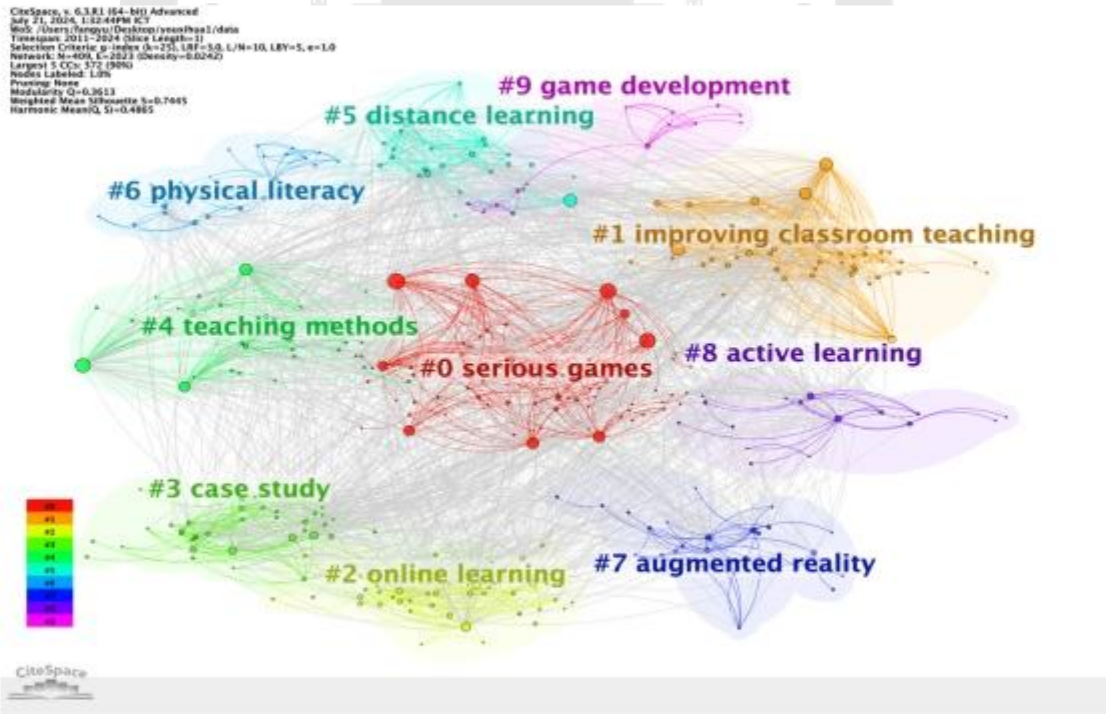


Figure 5 GL in the field of education research published keywords Clustering network knowledge graph

Table 2 GL in the field of education research keywords clustering

ID	Cluster Name	Silhouette	Mean (Year)	LLR
0	serious games	0.678	2017	serious games (23.31, 1.0E-4); game-based learning (17.95, 1.0E-4); higher education (14.51, 0.001); improving classroom teaching (5.99, 0.05); literature review (5.36,0.05)
1	improving classroom teaching	0.681	2019	improving classroom teaching (14.52, 0.001); intrinsic motivation (11.23, 0.001); learning analytics (10, 0.005); non-digital gamification (7.72, 0.01); students' motivation (7.32,0.01)
2	teaching and learning	0.679	2021	teaching and learning (12.87, 0.001); software process improvement (12.87, 0.001); problems and difficulties (12.87, 0.001); gamification (0.36, 1.0); higher education (0.19,1.0)
3	quantitative measurement	0.771	2020	quantitative measurement (12.36, 0.001); disaster exercises (12.36, 0.001); data gathering (12.36, 0.001); decision making (12.36, 0.001); higher education (0.24,1.0)
4	open badges	0.734	2018	open badges (12.36, 0.001); lifelong learning (12.36, 0.001); personal learning environment (12.36, 0.001); e-portfolio (12.36, 0.001); higher education (0.24,1.0)
5	spatial rehabilitation	0.707	2019	spatial rehabilitation (11.96, 0.001); research project (11.96, 0.001); recommendation system (11.96, 0.001); innovation (6.98, 0.01); learning (4.97,0.05)

Table 2 (Continue)

ID	Cluster Name	Silhouette	Mean (Year)	LLR
6	german (language)	0.876	2017	german (language) (11.96, 0.001); mobile educational services (I1.96, 0.001); second language (11.96, 0.001); foreign language (11.96, 0.001); computer application (11.96,0.001)
7	online learning	0.73	2020	online learning (18.2, 1.0E-4); educational innovation (17.26, 1.0E-4); active methodologies (13.44, 0.001); academic performance (12.9, 0.001); mobile learning (9.05,0.005)
8	case study	0.835	2018	case study (13.34, 0.001); educational technology (10.81, 0.005); game elements (10.38, 0.005); engagement (9.88,0.005); media literacies (9.07,0.005)
9	teaching methods	0.989	2016	teaching methods (8.36, 0.005); technology enhanced learning (8.36, 0.005); recommender systems (8.36, 0.005); educational technologies (8.36, 0.005); learning and teaching (8.36,0.005)
10	distance learning	1	2022	distance learning (11.18, 0.001); perceived usefulness (pu)(9.18, 0.005); online teaching(9.18, 0.005); technology acceptance (9.18, 0.005); massive open online courses(5.58,0.05)

Table 2 (Continue)

ID	Cluster Name	Silhouette	Mean (Year)	LLR
11	physical literacy	1	2015	physical literacy (7.56, 0.01); mhealth (7.56, 0.01); attentional bias modification (7.56, 0.01); game components (7.56, 0.01); attentional bias (7.56, 0.01)
12	augmented reality	1	2012	augmented reality (30.28, 1.0E-4); primary education (14.62, 0.001); stem education (13.48,0.001); gamification in education (13.19, 0.001); computational thinking (8.19,0.005
13	active learning	1	2013	active learning (23.67, 1.0E-4); blended learning (18.01, 1.0E-4); board game (13.02, 0.001); play (13.02, 0.001); escape room (9.28,0.005)
14	game development	0.991	2016	game development (18.38, 1.0E-4); pervasive games (18.38, 1.0E-4); games and learning (18.38, 1.0E-4); concurrent design (18.38, 1.0E-4); digital games (11.71, 0.001)

Note: LLR= Log-Likelihood Ratio

Upon further examination of the node information, the content of cluster groups, and their overlap in the map, it is observed that the research primarily focuses on three areas.

The first aspect focuses on teaching methods and student motivation, encompassing "improving classroom teaching," "teaching and learning," "teaching method," "active learning," and "distance learning." This category mainly explores methods to enhance intrinsic LM by improving classroom teaching. For instance, Silva (2022) and Zikas (2016) examine how augmented learning aids students in understanding complex concepts (Silva et al., 2022; Zikas et al., 2016). Blended learning, supplemented with gamification, compensates for the deficiencies of traditional teaching methods, enhancing the learning experience and outcomes (Nalyvaiko et al., 2021).

The second aspect concerns educational technology and innovation, divided into clusters such as "serious games," "open badges," "augmented reality," "online learning," "case study," and "game development." This category focuses on applying different educational technologies to GL. For example, "open badges" use reward-based badges to promote lifelong learning (Zainuddin et al., 2023). Augmented reality (AR) technology enhances human-computer interaction and learning experience through virtual simulations (Durão et al., 2020).

The third aspect involves the application of GL in specific fields, consisting of clusters like "quantitative measurement," "spatial rehabilitation," "German (language)," and "physical literacy." This category applies statistical methods and data analysis to quantify teaching effectiveness (Metwally et al., 2022). "Spatial rehabilitation" addresses the practical needs of special populations in medical education and rehabilitation, enhancing recovery outcomes in areas such as stroke, Parkinson's disease, and brain injuries (Barrett & Muzaffar, 2014; Holden, 2005; Koenig et al., 2009).

Discussion

GL is now widely applied across various educational stages, subjects.

In Primary Education, Ma Zong-bing (2018) conducted a semester-long teaching practice at an elementary school in Guangzhou, China, using gamified teaching strategies in programming classes to enhance students' computational thinking skills (Zong-bing, 2018). Du Yinuo explored how teachers should implement core literacy concepts into teaching, finding that gamified teaching in elementary mathematics enriched students' learning experiences in a joyful environment (Du, 2023). Jagušt et al. (2017) demonstrated through a case study in Croatian elementary schools that gamified digital courses significantly increased students' interest and motivation in learning (Jagušt et al., 2017).

In Higher Education, GL often employs points, badges, and leaderboards to boost student interest. Signori et al. (2018) found that these methods increased student engagement, motivation, and academic performance, and also promoted teamwork (Signori et al., 2018). Topîrceanu (2017) used role-playing to improve student attitudes toward learning and academic achievement (Topîrceanu, 2017). At a Swedish university, game elements like points and achievements were used to enhance self-directed learning and critical thinking in a course (Lindberg, 2019). Forndran and Zacharias (2019) reported high student acceptance and improved academic knowledge through gamified physics experiments (Forndran & Zacharias, 2019).

In Educational Training, GL is not only prominent in classroom settings but also in corporate project management and training. Armstrong et al. (2018) developed a roadmap for online training gamification using virtual rewards to enhance employees' understanding of their work (Armstrong & Landers, 2018). Koivisto (2022) applied GL in project portfolio management tasks, improving task allocation methods (Koivisto, 2022).

Conclusion

In just a few short years, GL has made significant theoretical and practical advances, yielding fruitful results in various domains. GL has demonstrated its efficacy across different educational stages and disciplines. By incorporating game design elements and blended learning models into classroom teaching, it has notably enhanced student motivation and improved learning outcomes. In the realm of modern educational technology innovation, artificial intelligence and learning analytics have provided personalized learning experiences, thereby enhancing teaching effectiveness. AR and VR technologies offer immersive learning experiences for students, while the application of gamified elements, such as open badges, promotes lifelong learning and career development.

Gamified Learning in Physical Education (GLiPE): Specialized Analysis

Bibliometric Analysis of GLiPE

To objectively describe the research frontiers of GLiPE, this study selects the Web of Science core collection database as the source of research literature. The search query used is TS= (gamification OR GL OR gamified education OR gamified training) AND TS= (PE OR sport education). The document types are limited to articles and conference papers, and the subject area is confined to educational research. The search covers the period from 2011 to 2024, yielding a total of 124 documents, which serve as the sample for this study.

Publication Trends

The number of publications is an important indicator of the development trend in a particular field over a specific period. It also provides an intuitive view of changes in research interest, which is crucial for analyzing the development trends of a field and predicting its future direction (Chen, 2006). This study uses Excel to analyze the sample literature

This study uses Excel to statistically analyze the sample literature. As shown in Figure 6, the earliest work is by Hebert et al. (2011), who researched dynamic virtual GL to enhance student reflection (Herbert et al., 2011). Before 2010, interest in GLiPE was relatively low, still in its nascent stage. Starting from 2014, the research entered a rapid development phase, with the number of publications increasing steadily. This surge is mainly attributed to the 2014 International Educational Technology Conference, which proposed using GL to improve student engagement and skills. The number of publications saw a sharp rise from 2018 onwards.

Although there was a slight decline in 2019, 2023 witnessed another peak. This resurgence could be due to the European Union's "Digital Education Action Plan" (Commission, 2020), which provided policy support for GL research (Yanli & Danni, 2021). Additionally, the COVID-19 pandemic, which led to reduced physical activity, psychological stress due to lockdowns, and isolation, also drove the application of GL in PE. Mendes et al. (2022) noted that gamification significantly advanced PE during the pandemic (de Sousa Mendes et al., 2022), corroborating this trend.

As of July 2024, eight articles on gamification in PE have been published. Overall, enthusiasm for research on GLiPE remains high, suggesting ample opportunities for further exploration.

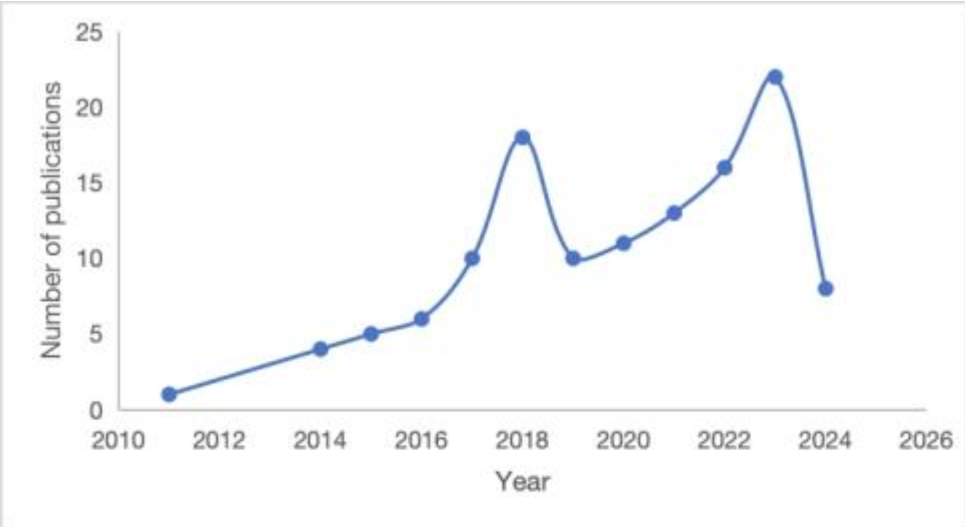


Figure 6 The annual publication of GLiPE

Analysis of Key Authors

In the co-authorship network map of a specific research field, authors with a high volume of publications indicate a significant influence within that field (Qiu et al., 2022). Additionally, the co-authorship network map reveals the collaborative relationships among different authors in the field (Pan & Zhang, 2023).

Table 3 shows that Fernández-Río, J. and Suárez Manzano, S. are the most prolific authors, each with four publications. Both researchers focus on the intersection of PE and GL, exploring how integrating gamification elements into PE can enhance student engagement, motivation, and learning outcomes. Other prolific authors include Kopniak, P., and Ferriz-Calero, A., who form the core group of researchers in the field of GLiPE, exerting significant influence in this research domain.

Using Price's law formula, $M = 0.749 \times \sqrt{Nmax}$ where *N* max is the number of publications by the most prolific author, which is 4 in this case), the threshold for prolific authors in the foreign sample is calculated to be 1.297. Therefore, authors with two or more publications are considered prolific in this study. The analysis reveals that prolific authors account for a total of 24 publications, representing 19.35% of the sample. This indicates relatively weak research collaboration among authors and a relatively low publication volume. While a few core authors and research teams have emerged, a large group of highly prolific and influential core authors and research teams has yet to form.

Table 3 Analysis of the important authors of GLiPE

Number of Publications (Articles)	Year	Author	Number of Publications (Articles)	Year	Author
3	2021	Fernandez-rio, J.	2	2021	Lopez serrano, S.
3	2021	Suarez manzano, S.	2	2022	Manzano carrasco, S.
2	2018	Kopniak, P.	2	2023	Bustamante, J.-C.
2	2022	Ferriz-valero, A.	2	2022	Perez-lopez, I.
2	2021	Jaen, J.	2	2023	Navarro-mateos, C.

Analysis of Key Institutions

Analyzing the collaboration among research institutions in a particular field helps quickly identify the main research forces and their distribution (Yin et al., 2024). Understanding the primary research institutions, core authors, and publication volumes in a field provides scholars with a clearer view of the current state and research strength both domestically and internationally (Wang et al., 2024). This study uses CiteSpace to process Chinese and international sample data, setting the node type to "institution" while keeping other settings as default. As shown in Table 4, the institutions with the highest number of publications are the University of Granada and the University of Zaragoza. These historic Spanish universities have made the most significant contributions to the field of GL in PE worldwide. Other notable institutions include the University of Oviedo, University of Sevilla, Universitat Politècnica de València, and Universidad de Jaén. The primary research institutions in this field are mainly universities, indicating a relatively uniform institutional type, with universities being the main driving force behind research on gamified teaching in PE.

Geographically, all research institutions focusing on GLiPE are concentrated in Spain. This concentration might be partly due to the 2017 publication of the "Common Digital Competence Framework for Teachers" in Spain (Cebrián-CiNotas et al., 2021), which catalyzed research activities. It also suggests that Spanish universities emphasize interdisciplinary collaboration, integrating knowledge from education, psychology, computer science, and other fields to conduct research on GLiPE.

The institution cooperation map shows 155 publishing institutions in the sample, with 87 connection lines between them, resulting in a cooperation network density of only 0.0073. This indicates that cooperation among research institutions is relatively low, highlighting the need for further improvement in research collaboration.

Table 4 Analysis of the important institutions of GLiPE

Number of Publications (Articles)	Institution	Nature	Location	Regional Attributes
4	University of Granada	University	Andalusia	Southern Spain
4	University of Zaragoza	University	Zaragoza	Northeastern Spain
3	University of Sevilla	University	Andalusia	Southern Spain
3	University of Oviedo	University	Asturias	Northwestern Spain
3	Universitat politecnica de Valencia	University	Valencia	Southeast Spain
3	Universidad de Jane	University	Andalusia	Southern Spain

Analysis of Research Hotspots in GLiPE

Keywords represent the core ideas of a paper and provide a concise summary of its theme. Analyzing the keywords in a specific field can help identify research hotspots. The higher the frequency of a keyword's occurrence, the greater the interest in that topic within the field. The size of a keyword node indicates its frequency:

larger nodes represent higher frequencies, while smaller nodes represent lower frequencies. This analysis helps to understand the dynamics and trends of the research field and to identify the main research directions and hotspots.

Using CiteSpace, this study processes the sample literature data by setting the node type to "keyword" and the threshold to 10, which means keywords that appear more than 10 times are selected. The resulting co-occurrence map of keywords in GL in PE includes 232 keyword nodes and 719 connections, with a density of 0.0268. The size of each keyword node represents its frequency; the larger the node, the higher its frequency.

After excluding eight keywords related to the search theme or similar terms ("gamification," " PE," "education," "physical activity," "games," "game-based learning," "educational game"), it is evident that "motivation" and "students" are among the most frequently occurring keywords. As shown in Figure 7, the varying diameters of the keyword nodes indicate significant differences in their frequencies. The largest nodes, such as "motivation" and "students," have the most connections, indicating that these keywords frequently appear in the same documents and serve as significant bridging terms.

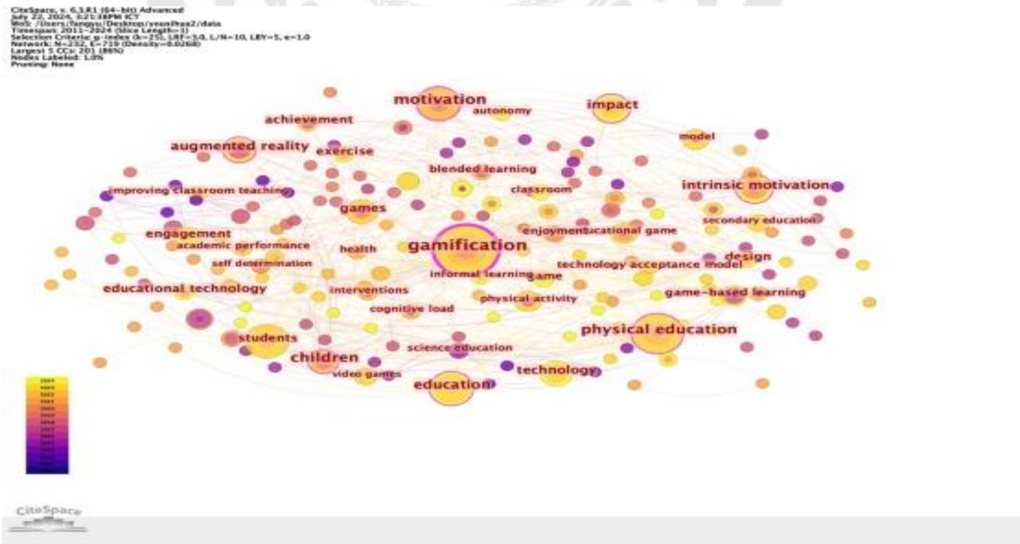


Figure 7 GLiPE research in the field of high frequency keywords co-Occurrence graph

The top 10 keywords by frequency are listed in Table 5. "Motivation" and "intrinsic motivation" rank first and second, indicating the crucial role of motivation in GL research. Studies show that gamification elements, through reward mechanisms and immediate feedback, can significantly enhance students' motivation and classroom engagement (Anderson & Dill, 2000; Ryan & Deci, 2017). The keyword "students" highlights the primary focus on the student population, showing that gamified strategies can significantly improve student participation and learning outcomes (Hamari et al., 2014). "Technology" and "augmented reality" reflect the application of modern technologies in PE, with AR technology notably enhancing students' learning experiences and interest (Ryan & Deci, 2017; Wu et al., 2013).

In summary, the research hotspots in GLiPE primarily focus on enhancing student motivation and the application of modern technologies to improve educational outcomes. These high-frequency keywords underscore the value of research in these areas, providing significant reference points for future educational research and practice.

Table 5 GLiPE research in the field of high frequency keywords statistics (frequency ≧10)

Item number	Keyword	frequency (times)	Centrality
1	motivation	12	0.19
2	students	11	0.02
3	intrinsic motivation	8	0.10
4	impact	8	0.10
5	children	7	0.18
6	technology	7	0.09
7	augmented reality	6	0.10
8	higher education	5	0.04
9	engagement	4	0.03
10	model	4	0.05

Due to the limited number of studies focusing on specific aspects, the Y value in CiteSpace was adjusted to 0.5 for burst keyword analysis. The resulting burst keywords (Figure 8) include "students" (2022-2024), " PE " (2022-2024), "education" (2022-2024), and "intrinsic motivation" (2020-2021). As shown in Figure 8, these keywords reflect shifts in research hotspots during their respective periods. The keywords with the highest burst strength are "students" and " PE," indicating that from 2022 to 2024, research on PE classroom teaching has garnered significant attention (Anderson & Dill, 2000).

Additionally, "intrinsic motivation" and "enjoyment" showed high burst strength from 2020 to 2021. This suggests that during the pandemic, despite the challenges posed by lockdown measures and remote learning, educators utilized various modern teaching tools (Winter et al., 2021). GL became a research hotspot, aiming to enhance students' intrinsic motivation and enjoyment in learning (Deci & Ryan, 2013).

Figure 8 GLiPE research mutation of keywords

Keyword Co-occurrence Cluster Analysis

To further explore the structural composition of research on GLiPE, a keyword co-occurrence cluster analysis was conducted using CiteSpace, as shown in Figure 9 and table 6. The modularity value (Q=0.5809) indicates a significant and well-defined cluster structure (Q > 0.3). The mean silhouette value (S=0.8934) suggests a high level of confidence in the clustering results (S > 0.5). This confirms the clarity and significance of the cluster structure, with distinct research themes within each cluster.

Keyword clustering groups similar keywords, aiding in the analysis of research hotspots and trend changes within the field. Further examination of the cluster group nodes and their contents reveals that the research primarily focuses on two main areas.

The first focus is on educational technology and learning methods. This category includes topics such as "science education," "architectures for educational technology systems," "higher education," "mobile learning," "active learning," "gamified m-learning," and "learning burnout." The primary exploration here is how innovations in educational technology and novel teaching methods can enhance educational outcomes. In terms of educational technology, Metrôlho et al. (2019) designed a web application to create and edit orienteering activities, which increased student interest in learning (Metrôlho et al., 2019). A Spanish scholar used the social network Edmodo to facilitate student communication, writing, and physical activities, showing a positive impact on student physical activity levels (Montiel-Ruiz et al., 2023). Additionally, an innovative health education program for Australian children aged 10-12, named "iEngage," was created to promote physical activity and behavioral changes, with results indicating significant increases in moderate and vigorous activities in the experimental group (Yacef et al., 2018).

Fernandez-Rio et al. (2020) applied non-modern technology-based gamification in the "science education" cluster, utilizing a blended teaching method. They divided students aged 6-14 into lower and higher grade levels and used MarvEF games as an intervention, which increased enthusiasm for PE. However, teachers expressed concerns about the increased workload associated with gamification (Fernandez-Rio et al., 2020). In the "higher education" cluster, Ferriz-Valero et al. (2020) recruited 127 university students to participate in various practical courses, including orienteering, rock climbing, water skills, and natural gymnastics. The study results indicated that GL improved academic performance, underscoring the effectiveness of gamification in higher education PE (Ferriz-Valero et al., 2020).

Secondly, the impact of education and its application to society focuses on "academic performance" and "design for social impact." This area examines the broader implications of educational applications on society. The COVID-19 pandemic in 2019 significantly reduced students' motivation to learn, increased the pressure of self-directed learning, and severely limited their academic performance (Grubic et al., 2020; Hammerstein et al., 2021). Prolonged absence from outdoor activities and irregular sleep patterns led to decreased aerobic capacity and a noticeable decline in PF among students (Guo et al., 2021; Lu et al., 2022). Consequently, online PE courses and AI-driven virtual reality applications emerged to address these challenges within this social context. Mujiono et al. (2021) reviewed literature and found that gamified online learning models can be effectively utilized in PE (Mujiono & Gazali, 2021). Additionally, various gamified applications such as "Kahoot!" "Nike+," and "Strava" have demonstrated the role of GLiPE, especially under the societal constraints imposed by the pandemic.

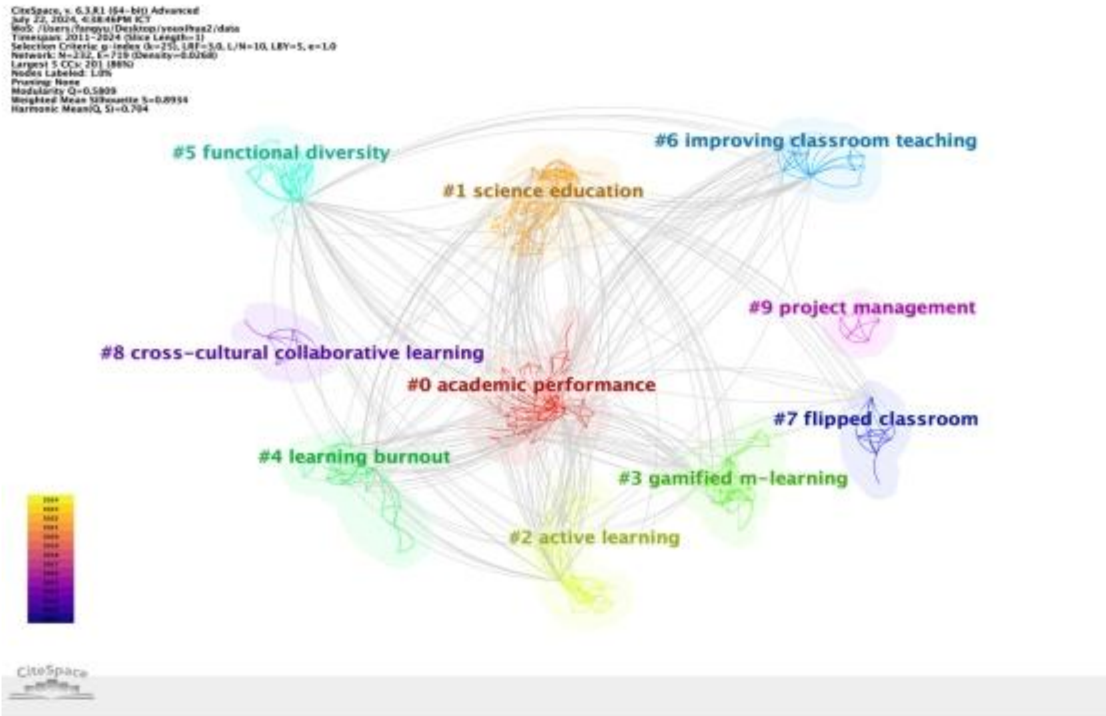


Figure 9 GLiPE keyword cluster map

Table 6 GLiPE keywords clustering statistics

ID	Cluster Name	Silhouette	Mean (Year)	LLR
0	academic performance	0.732	2018	academic performance (5.77, 0.05); motivation (4.3, 0.05); PE (3.64, 0.1); electronics (2.87, 0.1); java (2.87,0.1)
1	science education	0.755	2019	science education (7.51, 0.01); educational technology (4.02,0.05); game-based learning (4.02, 0.05); south africa (3.74,0.1); experiential education (3.74, 0.1)
2	architectures for educational technology system	0.791	2019	architectures for educational technology system (8.23, 0.005); cooperative/collaborative learning (8.23, 0.005); games (8.23,0.005); augmented and virtual reality (8.23, 0.005); mobile learning (4.47, 0.05)

Table 6 (Continue)

ID	Cluster Name	Silhouette	Mean (Year)	LLR
3	design for social impact	0.831	2019	design for social impact (8.23, 0.005); covid-19 pandemic (8.23, 0.005); gamified apps (8.23, 0.005); app design (8.23,0.005); gamification (0.57,0.5)
4	higher education	0.797	2019	higher education (7.49, 0.01); proactive learning (6.35, 0.05); mixed reality (6.35, 0.05); mobile learning technology (6.35, 0.05); student expectations (6.35,0.05)
5	Mobile learning	0.758	2019	Mobile learning (13.45,0.001); meta-verse (13.45,0.001); augmented reality (13.19,0.001); virtual reality (9.07,0.005); online teaching (8.95,0.005)
6	active learning	0.872	2020	active learning (9,0.005); educational innovation (9,0.005); significant learning (4.47, 0.05); multi-display environments (mde) (4.47, 0.05); neuroscience (4.47,0.05)
7	gamified m-learning	0.925	2016	gamified m-learning (5.29, 0.05); extended unified theory of acceptance and use of technology (5.29, 0.05); train the trainer (5.29, 0.05); people (5.29, 0.05); prevention program (5.29,0.05)
8	learning burnout	1	2016	learning burnout (5.1, 0.05); soft skills (5.1, 0.05); students (5.1, 0.05); motivation programming (5.1, 0.05); mesomeric effect (5.1, 0.05)

Evolution Path Analysis

Evolutionary path analysis allows subsequent researchers to clearly understand the development direction of research hotspots and predict future research trends (Mi et al., 2024; Shen et al., 2022). We use the Timeview function of CiteSpace to visualize the evolutionary path by analyzing the chronological order of publications. Additionally, we perform thematic clustering of relevant keywords (Figures 10 and 11) (Wang & Lu, 2020).

Preliminary Development Stage (2011-2015). Research during this period primarily explores the implementation of GL to enhance student motivation, utilizing methods such as points-based rewards and leaderboards. With advancements in science and technology, "augmented reality" and "virtual reality" (Bower et al., 2014; Wu et al., 2013) began to be introduced into PE, entering the research scope of experts and scholars.

Rapid Development Period (2016-2020). Keywords such as "physical literacy," "augmented reality," and "health education" further demonstrate technological advancements in the field of PE. Moreno-Guerrero et al. (2020) conducted a study that trained students' spatial orientation skills using AR technology, proving the effectiveness of this training method (Moreno-Guerrero et al., 2020). Additionally, two other studies found that AR-based teaching was more effective than video-assisted teaching, particularly for learning complex motor skills (Chang et al., 2020; Silva et al., 2022).

Mature Development Phase (2021-Present): As research continues to become more refined and detailed, terms like "higher education," "mobile learning," and "cognitive engagement" have emerged as key focus areas. Studies by Flores-Aguilar, G. (2023) and Fernández-Río, J. (2020) highlight the practical significance of GL, showing a positive correlation with enhancing students' motivation and PF. Yu et al. (2018) demonstrate that mobile applications in PE assist students in effectively implementing their plans (Yu et al., 2018).

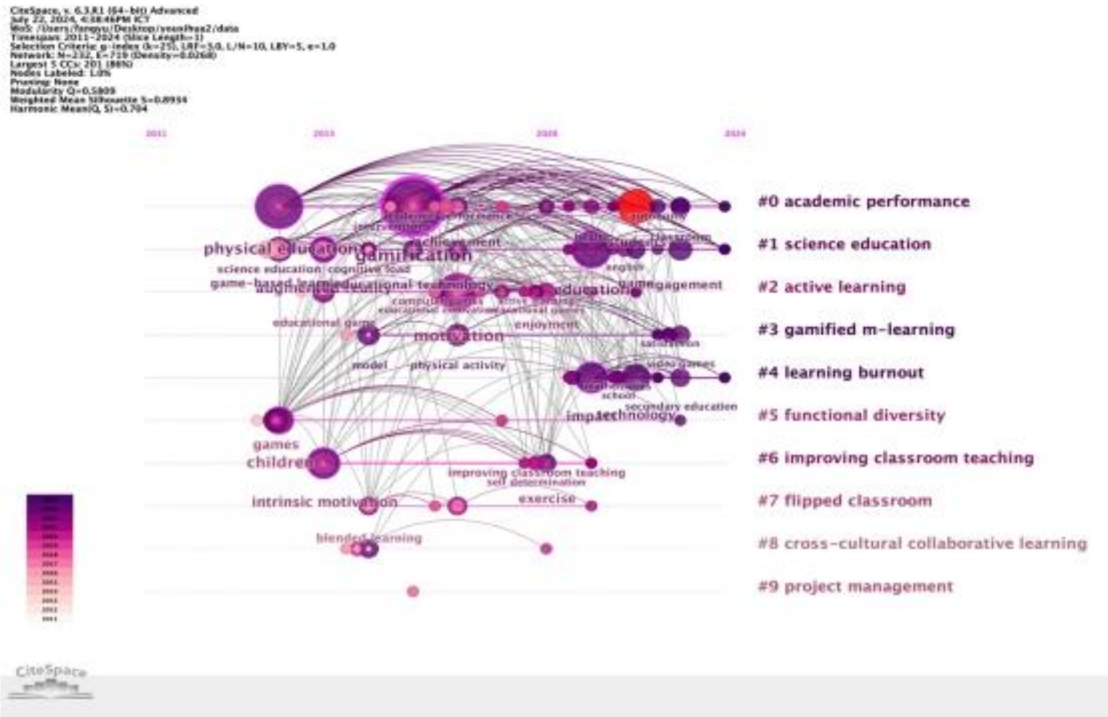


Figure 10 The evolution path diagram of GMiPE

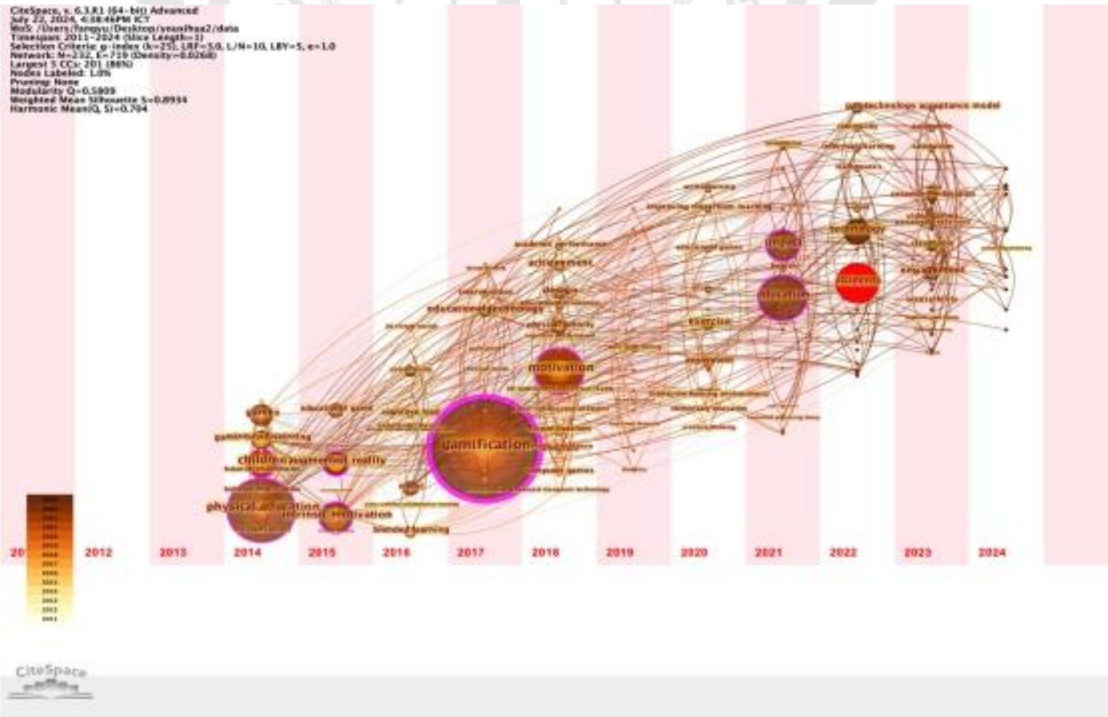


Figure 11 GMiPE research keywords sequence map

Discussion

We explore the current research on GLiPE from the following three perspectives.

The first aspect, the application of non-modern educational technology in GLiPE within PE. This category mainly features live games implemented in PE classes, prominently represented by Spanish scholars such as Fernández-Río, J., Flores-Aguilar, G., and Alberto Ferriz-Valero. These researchers investigate how incorporating live games in PE can enhance students' motivation and academic performance, yielding significant results. They also note improvements in classroom interaction and collaboration. However, some scholars highlight that designing gamified courses can be time-consuming for teachers, increasing their workload.

Application of Modern Educational Technologies in PE With the development of AR and VR technologies, educational technology has significantly advanced. AR technology, through immersive learning modes, enhances student experiences and helps simulate complex physical movements, aiding in better understanding. This technology generates a high level of student interest in learning. Teachers can also use big data to track students' learning progress, adjusting the pace and difficulty of instruction as needed Students can personalize their learning experiences through applications, achieving effective human-computer interaction.

Application of GL in PE within a Social Context. Modern technology is constantly evolving, and the COVID-19 pandemic in 2019 served as a catalyst for significant advancements in educational technology. The constraints of offline learning have expanded the opportunities for online distance learning and the integration of AR technology, effectively addressing students' urgent need to alleviate learning anxiety. This societal backdrop has driven educational progress, particularly by relieving the psychological stress of students through online PE courses, thereby promoting MH. However, related studies indicate that while online courses have somewhat improved students' PF, their level of outdoor activity and aerobic capacity remain relatively low. Therefore, when selecting teaching methods, it is crucial to consider the social environment to choose the most appropriate methods that further students' development. This is a significant issue that warrants careful consideration and in-depth exploration.

Conclusion

There is no doubt that GL has significantly improved students' motivation and efficiency in PE. Its application in non-modern educational technologies has also shown remarkable results. AR and VR technologies have further enhanced learning outcomes, particularly in understanding complex technical movements. Additionally, the development of big data and AI technologies provides students with more personalized learning opportunities and enables teachers to monitor student progress in real time.

However, dialectical thinking reminds us to consider both sides of an issue. The advancement of technology is a double-edged sword. When selecting modern teaching technologies, we must consider practical circumstances and actual needs rather than blindly following trends. It is crucial to choose appropriate gamified educational technologies to promote the development of PE.

Application of SDT in PE

SDT is a crucial framework for understanding motivation (R. M. Ryan & E. L. Deci, 2000). It primarily explores how to fulfill three basic psychological needs: autonomy, competence, and relatedness. This discussion will focus on the application of these three needs in PE at vocational colleges, providing theoretical foundation for the current study's intervention design.

A systematic review employing structural equation modeling analyzed data from 265 studies, supporting that fulfilling SDT's needs for autonomy, competence, and relatedness is closely linked to students' intrinsic motivation (Vasconcellos et al., 2020). One study applied SDT through a "SPOC+flipped classroom" model in a PE setting, conducting experiments with sports major students. The experimental group received more support for their needs, leading to higher need satisfaction, promoting internal motivation, and maintaining high levels of autonomous motivation (Hu et al., 2022).

SDT and Expected Differential Effects: SDT effectively promotes students' basic psychological needs. However, Bergh et al. (2014) integrated 74 studies on SDT, emphasizing the need for future research to combine teaching and psychology further to explore how SDT can enhance student engagement in PE (Van den Berghe et al., 2014). The theory predicts that interventions directly satisfying psychological needs (motivation outcomes) will show larger effects than those requiring behavioral mediation (PF) or addressing complex psychological states (clinical MH symptoms).

PF, MH, and Motivation in Vocational Students

Current Status and Challenges

PF Status

Studies indicate that the overall PF levels of vocational college students are relatively low (Shen et al., 2021; Waldron & Dieser, 2010). Research also indicates that students in economically developed regions benefit from better sports facilities, higher levels of parental support, and well-structured PE programs. These factors provide more opportunities for exercise, thereby enhancing their PF levels (Gan et al., 2019). Gender differences in students' PF levels are also evident. Generally, male students tend to excel in muscle strength and cardiovascular endurance, whereas female students perform better in flexibility tests (Pampel et al., 2010).

MH Issues

Common MH issues among students in HVC include anxiety, depression, and academic stress (Pedrelli et al., 2015). A survey using the SCL-90 to assess MH issues among Chinese vocational college students found a prevalence rate of 12.22% for psychological problems (Hong-mei, 2006). Another study focusing on HVC students discovered that the prevalence of depression is approximately 15-25% (Lamlé et al., 2023). Additionally, related research has explored the relationship between academic pressure, depression, and anxiety. For example, Zhang Chang et al. (2022) conducted a study involving 1,309 students at a Chinese school, revealing a positive correlation between academic pressure and levels of depression and anxiety (Zhang et al., 2022)).

LM Factors

Student motivation is a key factor influencing their learning outcomes and engagement (Esra & Sevilen, 2021). SDT provides a theoretical framework for understanding and enhancing student motivation (R. M. Ryan & E. L. Deci, 2000). When students have the autonomy to choose the type and manner of physical activities, their intrinsic motivation is significantly enhanced. By setting appropriately challenging tasks and providing immediate feedback, students' sense of competence can be increased. Additionally, positive social interactions and teamwork can enhance students' sense of relatedness, improving their feelings of belonging and support during the learning process.

The Relationship Between Physical Activity and MH

The role of physical activity in improving MH has been widely researched and acknowledged. Understanding how physical activity impacts students' MH in the context of HVE is crucial for developing effective intervention measures.

Research indicates that regular participation in physical activity can significantly reduce symptoms of anxiety and depression (Rebar et al., 2015). Engaging in physical activities promotes the release of endorphins in the body, which helps to enhance mood and reduce stress (Adling & Chamle, 2021; Ali et al., 2021). A study found that students who participate in physical activities experience a significant reduction in anxiety and depression symptoms. Moreover, physical activity also helps enhance students' self-esteem and self-efficacy (Jensen et al., 2022).

Many studies indicate that regular physical activity is closely associated with improved sleep quality (Wang & Boros, 2021). Good sleep quality is crucial for maintaining MH (Scott et al., 2021), as it helps students better cope with daily stress and challenges.

Current State of PE Teaching in HVC

Curriculum Content and Teaching Methods

Studies indicate that the curriculum content in Chinese HVC is often monotonous. Approximately 75% of students find the course content unappealing, and

45% exhibit low motivation to learn (Wang, 2014). A pervasive issue in the PE curriculum is the lack of engaging and progressive teaching philosophies (Zhang, 2013). Additionally, research highlights that the curriculum often fails to align with vocational needs, neglecting the physical demands of specific careers (Xue & Guo, 2024; Zhang, 2013).

Many students feel that the current teaching methods in vocational college PE are dull and lack interactivity, with teachers primarily lecturing without engaging students (Zhang, 2022). Scholars argue that, despite the advances in information technology, the teaching methods remain outdated and fail to meet the developmental needs of students and societal demands (Lin, 2021).

In summary, there are numerous challenges in the current state of vocational college PE teaching. Scholars are actively exploring ways to improve these conditions. Some suggest focusing on practical skills and aligning the curriculum with talent cultivation plans (Jun, 2022). Others advocate for integrating PE with medical education (Niu, 2023), keeping pace with the information age, and incorporating the Internet of Things and artificial intelligence into the curriculum (Yu & Mi, 2023).

Research Gaps and Study Positioning

Limited Rigorous Experimental Designs in Vocational Contexts

Most existing gamification research in vocational education relies on pre-post designs without adequate control groups or randomization. The current study addresses this limitation through Solomon four-group experimental design that controls for testing effects, selection bias, and historical factors while specifically focusing on HVE contexts.

Absence of Multi-Domain Outcome Assessment

Previous research typically examines single outcome domains (motivation OR performance OR health) rather than comprehensive assessment across multiple domains. The current study's simultaneous examination of PF, MH, and LM addresses this gap while providing insights into differential effect patterns predicted by SDT.

Lack of Mechanism Investigation

Few studies investigate the mediating processes through which gamification produces its effects. The current study's examination of game performance as a potential mediator between psychological benefits and sustained engagement fills this theoretical gap while testing specific pathways predicted by SDT and game mechanics theory.

Synthesis and Theoretical Model

The literature review supports a theoretical model where GL produces differential effects across outcome domains through distinct mechanisms:

Direct Effects

LM shows largest effects through immediate psychological need satisfaction via autonomy support, competence building, and relatedness enhancement.

Mediated Effects

PF improvements occur through behavioral mediation requiring sustained engagement over time, with effects concentrated in performance-based measures rather than body composition.

Selective Effects

MH benefits concentrate on environmental stress reduction rather than clinical symptom amelioration, reflecting the situational nature of gamified interventions.

This differential effect pattern, predicted by SDT and supported by emerging empirical evidence, forms the theoretical foundation for the current study's hypotheses and provides the conceptual framework for interpreting anticipated results. The integration of game mechanics theory and constructivist learning principles further supports the expectation that gamified approaches will be particularly effective in vocational education contexts where practical skill development and professional preparation are primary objectives.

CHAPTER 3
METHODOLOGY

Research Design

Use of Solomon Four-Group Design

This study employs a randomized controlled trial using the Solomon four-group design to evaluate the effectiveness of gamified learning interventions in higher vocational PE. The Solomon four-group design was first proposed by American psychologist R. L. Solomon in 1949 to investigate the impact of specific factors on participants' behavior while controlling for multiple threats to internal validity (Solomon, 1949).

Benefits of Solomon Four-Group Design

Control for Pretest Effects: The Solomon design eliminates potential sensitization effects that may occur when participants are exposed to pretesting. By including groups without pretesting (EG2 and CG2), the design can detect whether the pretest itself influences post-intervention outcomes, ensuring that observed changes are attributable to the intervention rather than testing artifacts (Braver & Braver, 1988). This is particularly important in educational research where pretesting may alert participants to study objectives and modify their subsequent behaviors (Mai et al., 2020).

Reduction of Response Bias: The design reduces multiple sources of bias including social desirability bias, expectancy effects, and demand characteristics. Participants in no-pretest groups are less likely to modify their responses based on perceived researcher expectations, as they lack baseline comparison points that might influence their post-test responses (Kirisci et al., 2020).

Enhanced Internal Validity: By controlling for selection bias, maturation effects, historical influences, and testing effects simultaneously, the Solomon design provides superior internal validity compared to conventional pre-post designs. This is particularly crucial in educational interventions where multiple factors may influence outcomes over time (Lavrakas, 2008).

Statistical Power and Effect Size Precision: The four-group structure enables more precise estimation of treatment effects while controlling for various threats to validity. This design allows for detection of both main effects and interaction effects between testing and treatment conditions (Braver & Braver, 1988).

The Solomon four-group experimental design is employed to control confounding factors while identifying causal effects, thereby enhancing the reliability and validity of the experiment (Kirisci et al., 2020). This design is widely applied in educational research and is recognized as an effective approach for establishing causal relationships in complex educational interventions.

Participants in the Study

Target Population

The target population consists of first-year students enrolled in HVE programs at health-oriented institutions in China. Specifically, the study focuses on students aged 18-25 years enrolled in nursing, rehabilitation, and pharmacy programs at Nanchang Health Vocational and Technical College during the 2024 academic year.

Sample Size Calculation

Power Analysis Using G*Power: The sample size calculation was conducted using G*Power 3.1 software to ensure adequate statistical power for detecting meaningful intervention effects (Cohen, 1988; Faul et al., 2007). The calculation was based on a 2×2 mixed ANOVA design examining main effects and interaction effects between treatment and testing conditions (Kirisci et al., 2020; Stanley & Campbell, 1963).

The statistical parameters for the power analysis are determined in accordance with established standards in educational and psychological research. A medium effect size (f = 0.25) is adopted, consistent with Cohen's guidelines and supported by interpretations identifying f = 0.25 as a medium effect (Cohen, 1988). The alpha level is set at 0.05, reflecting conventional significance thresholds (Lakens, 2013) and widely accepted practice in educational studies (Fewtrell et al., 2008). Statistical power is specified at 0.80, in line with recommendations by Cohen (Cohen, 1992) and supported by methodological texts in education (Maxwell & Delaney, 2003). The study employs a Solomon four-group design to enhance causal inference (Kirisci et al., 2020; Stanley & Campbell, 1963), with two measurements (pre- and post-test) applied where appropriate (Dimitrov & Rumrill, 2003).

Power Analysis Results: G*Power analysis indicated a minimum sample size of 45 participants per group for detecting medium effect sizes with 80% power (Cohen, 1992). To account for potential attrition (estimated at 10% based on similar educational intervention studies) (Dettori, 2011; Fewtrell et al., 2008), the target sample size was set At least 50 participants per group, resulting in a total sample of 200 participants.

Effect Size Justification: The medium effect size (f = 0.25) was selected based on meta-analytic evidence from gamification studies in educational contexts, which typically report effect sizes ranging from 0.20 to 0.40 for motivational outcomes (Hattie, 2008; Sailer & Homner, 2020). This conservative estimate ensures adequate power while accounting for potential variability in vocational education settings (Cohen, 1988).

Participant Selection and Recruitment

Initial Assessment and Recruitment: A total of 220 students were initially assessed for eligibility through the "Wenjuanxing" online platform. Registration information was disseminated through WeChat class groups to ensure broad accessibility across target programs (nursing, rehabilitation, and pharmacy).

Screening and Randomization: After applying inclusion and exclusion criteria, as shown in Table 7 and Figure 12, 200 eligible participants were randomly assigned to one of four groups using computer-generated randomization sequences created in Excel. Block randomization with varying block sizes (4, 6, 8) was employed to ensure balanced group allocation throughout the recruitment period while maintaining allocation concealment (Schulz et al., 2011).

Table 7 Participant Flow and Retention Summary

Phase	EG1	CG1	EG3	CG2	Total
Randomized	50	50	50	50	200
Completed Intervention	48	49	50	50	197
Analyzed	48	49	50	50	197
Retention Rate	96%	98%	100%	100%	98.50%

Inclusion and Exclusion Criteria

Inclusion Criteria

Participants are eligible if they are first-year students aged 18 to 25 during the 2024 academic year, enrolled in nursing, rehabilitation therapy, or pharmacy programs at Nanchang Health Vocational and Technical College. Eligibility further requires voluntary participation with signed informed consent, completion of the Physical Activity Readiness Questionnaire (PAR-Q) with all responses marked "NO," the ability to attend the full 16-week intervention period regularly, and sufficient Chinese literacy to complete all questionnaires independently.

Exclusion Criteria

Participants are excluded if they have prior formal experience with GL environments or structured educational gaming programs, are currently undergoing treatment for severe MH conditions requiring professional psychological intervention (e.g., major depression, severe anxiety, or psychotic disorders), or are under 18 years of age without parental or guardian consent. Exclusion also applies to individuals identified by the PAR-Q as having medical contraindications to physical activity, those unable to commit to the full 16-week intervention due to academic or personal conflicts, and students with physical disabilities that prevent safe participation in the prescribed PE activities.

Participant Screening and Exclusion Process

As part of the initial eligibility assessment, the PAR-Q (See Appendix A) is administered to ensure participant safety during physical activity. This screening tool is recognized for its reliability and validity in identifying individuals at potential risk when engaging in exercise (Shephard, 1988). Among the 220 students assessed, 20 are excluded based on the following criteria: 15 students provide one or more "YES" responses on the PAR-Q, indicating possible health risks; 3 students are under the age of 18; and 2 students withdraw participation after declining to sign the informed consent form despite receiving full study information.

Informed Consent Process

The study was conducted among first-year students from Nanchang Health Vocational and Technical College in Jiangxi Province, China. The remaining 200 eligible students proceeded to sign informed consent forms, indicating their voluntary participation and understanding of the study's purpose, procedures, potential risks, and benefits, following ethical guidelines (Goldstein et al., 2018). The consent process ensured that all participants were fully informed about the Solomon four-group design, randomization procedures, and their right to withdraw from the study at any time without consequences (see Appendix B-1 and Appendix B-2).

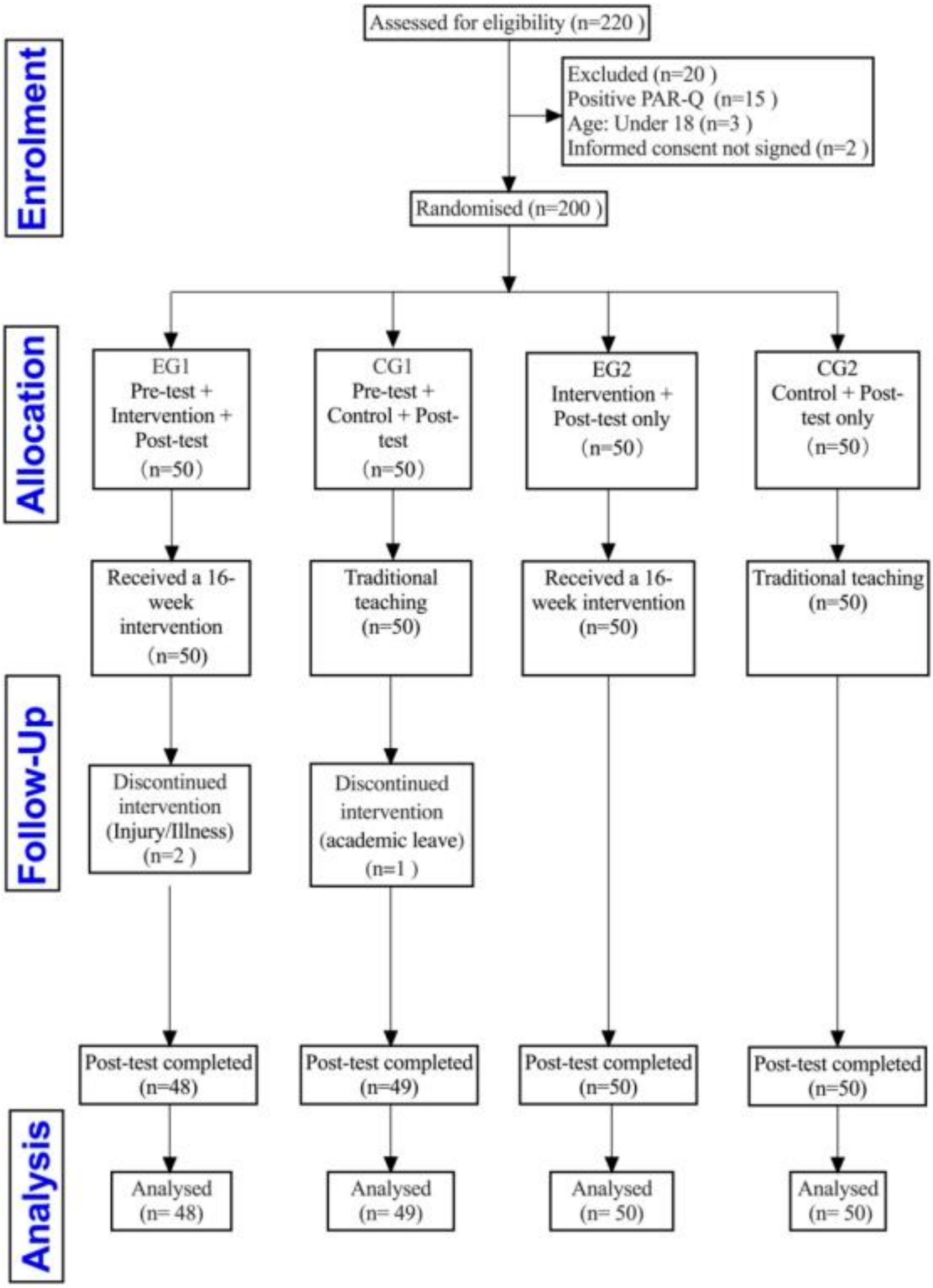


Figure 12 CONSORT Flow Diagram for Solomon Four-Group Design

Ethical Considerations

Ethical approval was obtained from Nanchang Health Vocational and Technical College prior to participant recruitment and data collection (Approval No. NCHVC-2026RT-25-6) (see Appendix C).

All participants provided informed consent after receiving comprehensive information about study purpose, procedures, risks, and benefits. The consent process included explanation of the Solomon design, randomization procedures, and the voluntary nature of participation with right to withdraw without consequences.

While physical activity risks were minimal, all participants completed PAR-Q screening to identify contraindications. Safety protocols included trained personnel availability and emergency procedures during physical assessments.

All data were coded with participant ID numbers, with identifying information stored separately in password-protected files. PF data were managed by the school's training center and stored in the internal school database, while psychological assessment and LM data were overseen by the school's MH center to ensure data confidentiality and prevent any leakage.

Research Instruments

The selection of appropriate research tools is crucial to the success of scientific studies (Moloney & Maple, 2023). The use of correct tools enhances research efficiency, minimizes errors, and ensures reproducibility and reliability. Therefore, choosing suitable research tools is not only a technical requirement but also a necessary step to maintain scientific rigor. This study employs three primary assessment instruments to measure PF, MH, and LM, each selected based on established psychometric properties and appropriateness for the target population.

Instrument Selection Rationale and Original Psychometric Properties

PF Assessment: Chinese National Student PF Standard

For PF assessment, this study adopts the National Student PF Standard (Revised 2014) issued by the Ministry of Education of China. This standardized assessment system was selected for several evidence-based reasons. First, it serves as the official national guideline for evaluating PF across all educational levels in China, providing a systematic, comprehensive, and objective assessment framework (China, 2014). The standard comprehensively assesses students' physical health levels based on body shape, body function, and PF, forming an important component of China's national core literacy system and academic quality standards for student development.

Second, extensive validation studies have demonstrated the reliability and validity of this testing system. Research by Yi et al. (2019) confirmed that the Chinese National Student PF Standard battery is a reliable and valid instrument to assess PF in adolescents, with test-retest reliability coefficients (ICC) exceeding 0.90 across all assessments, which is considered excellent (Yi et al., 2019). Similarly, Wang et al. (2012) validated this testing system with large-scale samples across multiple provinces, demonstrating its robust psychometric properties for Chinese youth populations (Wang et al., 2012).

Third, the standardized scoring system aligns with China's educational policy framework. The 2014 revision specifically enhanced the reliability, validity, and discriminative power of the Standard's application, strengthening its functions of motivating students, providing feedback for adjustment, and guiding physical exercise (China, 2014). The scoring guidelines provide age- and sex-specific norms based on extensive national surveillance data, ensuring culturally appropriate benchmarks for Chinese vocational students.

The testing battery includes seven components: height and weight (body composition), vital capacity (cardiopulmonary function), sit-and-reach (flexibility), standing long jump (explosive power), sit-ups for females/pull-ups for males (muscular strength and endurance), 50-meter sprint (speed), and 800-meter run for females/1000-meter run for males (cardiorespiratory endurance). These components comprehensively assess multiple dimensions of health-related PF relevant to this study's objectives.

MH Assessment: Depression Anxiety Stress Scales-21 (DASS-21)

The DASS-21 was selected to assess MH outcomes based on its well-established psychometric properties and cross-cultural validity. Originally developed by

Lovibond and Lovibond (1995), the DASS-21 is a 21-item self-report instrument designed to measure three related negative emotional states: depression, anxiety, and stress (Lovibond, 1995). The scale has demonstrated robust psychometric properties across diverse populations and cultural contexts.

Original psychometric validation studies have established strong reliability for the DASS-21. Henry and Crawford (2005) reported internal consistency reliability (Cronbach's α) of 0.93 for the total scale, with subscale reliabilities of 0.88 for depression, 0.82 for anxiety, and 0.90 for stress in a large non-clinical adult sample (n=1,794) (Henry & Crawford, 2005). Antony et al. (1998) further confirmed these findings across clinical and community samples, reporting internal consistency coefficients ranging from 0.87 to 0.97 for the DASS-21 subscales, demonstrating excellent reliability in both clinical groups and community samples (Antony et al., 1998).

Confirmatory factor analysis has consistently supported the three-factor structure of the DASS-21. Henry and Crawford (2005) demonstrated that a bifactor model—comprising three specific factors (depression, anxiety, stress) plus a general psychological distress factor—provided the best fit to the data (χ^2=893.7, df=162, CFI=0.94, SRMR=0.03, RMSEA=0.05), indicating strong structural validity (Henry & Crawford, 2005). Brown et al. (1997) replicated this three-factor structure in clinical populations, with exploratory factor analysis accounting for 55% of item variance and confirmatory factor analysis supporting distinct but correlated depression, anxiety, and stress factors (Brown et al., 1997).

The DASS-21 has demonstrated excellent cross-cultural validity, with successful validations in Asian populations particularly relevant to this study. Multiple studies have confirmed the instrument's psychometric integrity across diverse cultural contexts, supporting its appropriateness for Chinese vocational student populations (Norton, 2007; Oei et al., 2013). The scale's focus on symptoms rather than diagnoses, combined with its brevity (21 items completed in approximately 5 minutes), makes it particularly suitable for large-scale educational research contexts.

LM Assessment: Physical Activity and Leisure Motivation Scale (PALMS)

The PALMS was selected to assess motivation for physical activity participation based on its comprehensive theoretical foundation and established psychometric properties. Developed by Morris and Rogers (2004) as a shortened version of the 73-item Recreational Exercise Motivation Measure (REMM), the PALMS is a 40-item instrument measuring eight distinct motivational dimensions: mastery, enjoyment, psychological condition, physical condition, appearance, others' expectations, affiliation, and competition/ego (Morris & Rogers, 2004).

The development of PALMS followed a rigorous empirical and theoretical approach. Morris and Rogers (2004) conducted extensive item analysis—examining means, standard deviations, skewness, kurtosis, factor loadings, item-subscale correlations, and item-deleted alpha coefficients—to select the five strongest items from each of the eight REMM factors, creating a more efficient 40-item measure while maintaining robust psychometric properties (Morris & Rogers, 2004).

Original validation studies have established strong reliability for PALMS across diverse populations. Molanorouzi et al. (2014) validated PALMS with 309 Malaysian participants engaged in various physical activities, reporting internal consistency coefficients (Cronbach's α) ranging from 0.80 to 0.99 across the eight subscales, demonstrating excellent reliability (Molanorouzi et al., 2014). Test-retest reliability over two weeks yielded correlation coefficients ranging from 0.76 to 0.91 ($p<0.01$), indicating strong temporal stability (Molanorouzi et al., 2014).

Confirmatory factor analysis has supported PALMS' eight-factor structure across multiple validation studies. Chowdhury's (2012) Australian validation study (n=202) reported robust factor structure with excellent model fit indices: CMIN/DF=2.22, NFI=0.95, CFI=0.97, RMSEA=0.078 (Roychowdhury, 2012). Zach et al. (2012) validated the Hebrew version (PALMS-H) with 678 Israeli participants aged 9-89 years, demonstrating good internal consistency (α=0.63-0.96) and supporting the instrument's applicability across diverse age groups (Zach et al., 2012).

The PALMS' theoretical grounding in SDT makes it particularly appropriate for this gamification study. The eight subscales can be categorized according to SDT's intrinsic-extrinsic motivation continuum: intrinsic factors (mastery, enjoyment) and extrinsic factors (physical condition, psychological condition, appearance, others' expectations, affiliation, competition/ego). This alignment enables examination of how gamified learning differentially affects various motivational orientations, directly addressing the study's theoretical framework and research questions.

Cross-cultural validation studies have demonstrated PALMS' reliability across Asian contexts. Van Lankveld et al. (2021) validated the Dutch version with excellent psychometric properties, while validated versions exist in Persian (Molanorouzi et al., 2014), Malay (Kueh et al., 2018), and Spanish (multiple versions), supporting its cross-cultural applicability (Kueh et al., 2018; van Lankveld et al., 2021; Zarei et al., 2016). The availability of these validated Asian versions provides confidence in the instrument's appropriateness for Chinese vocational student populations.

Gamification Scoring System

The gamified learning intervention employs a comprehensive scoring system designed to quantify student participation and performance throughout the 16-week course. This mechanism serves both motivational and analytical functions, providing participants with immediate feedback while generating quantitative data for research analysis. The system is based on a football league-style point allocation mechanism, where students accumulate points through participation and achievements in eight distinct gamified activities (see Appendix G for detailed activity designs), conducted twice per week, with activities from weeks 1-8 repeated beginning in week 9.

Point Allocation Criteria: Non-participation is assigned 0 points, while active participation is awarded 1 point per activity. A winning team receives an additional 3 points, making the maximum attainable score per activity 4 points (1 for participation plus 3 for victory). Over the 16-week intervention period (32 total sessions), students can accumulate a maximum total score of 128 points, with individual activity scores ranging from 0 to 16 points per activity type.

Psychometric Properties of Research Instruments

Prior to examining the intervention effects, the psychometric properties of all measurement instruments were evaluated to ensure data quality and analytical validity.

PF Testing: Reliability and Validity Evidence

Reliability and Validity Assessment

To establish the psychometric properties of the PF testing instruments, both test-retest reliability and content validity were assessed prior to the main study.

Participants: A random sample of 331 students from the 2023 cohort of vocational college programs was recruited during the annual PF testing period (34 males, 297 females; mean age = 18.4 ± 0.73 years). All participants provided informed consent.

Test-Retest Reliability Procedure: The complete fitness battery was administered twice, separated by a two-week interval to allow adequate recovery while minimizing maturation effects. Testing conditions remained consistent across both time points. Indoor assessments (height/weight, vital capacity, sit-and-reach, standing long jump, sit-ups/pull-ups) were conducted at the Indoor PF Testing Center; outdoor assessments (50-meter sprint, 800m/1000m runs) were performed on the athletic track (see Appendix D, Figures D1-4).

Content Validity Assessment: Five content experts evaluated the representativeness and relevance of the testing battery. Experts were directors of physical education departments at five vocational colleges in Jiangxi Province, selected based on their dual roles as administrators and frontline PE instructors, ensuring both theoretical knowledge and practical expertise in vocational student fitness assessment. A standardized Content Validity Assessment Form (Appendix I) employed a 5-point relevance scale (1=not relevant, 5=highly relevant), with scores of 4-5 indicating high representativeness, relevance, and importance (Polit & Beck, 2006; Polit et al., 2007).

DASS-21: Reliability and Validity Evidence

The DASS-21 is a 21-item self-report instrument designed to measure three negative emotional states: depression (7 items), anxiety (7 items), and stress (7 items). Each item is rated on a 4-point scale (0 = did not apply at all, 3 = applied very much or most of the time), with subscale scores ranging from 0 to 21. Higher scores indicate greater severity of symptoms. Original validation studies demonstrate excellent psychometric properties, including strong internal consistency (α = 0.82–0.93), robust three-factor structure, and cross-cultural validity across diverse populations (see Chapter 3 for detailed original psychometric evidence and selection rationale).

Reliability and Validity Assessment

Following standard guidelines recommending 5-10 participants per item (Lee, 2007), and with the DASS-21 containing 21 items, a minimum sample of 105-210 participants was required. In August 2024, surveys were administered via the "Wenjuanxing" online platform to students at four vocational colleges in Jiangxi Province. Of 800 distributed questionnaires, 732 valid responses were obtained (91.5% response rate), exceeding the minimum requirement. The sample comprised 263 males (35.9%) and 469 females (64.1%), distributed across first-year (n = 288, 39.3%), second-year (n = 383, 52.3%), and third-year (n = 61, 8.3%) students (see Appendix E for complete questionnaire).

Statistical Analyses: Data were analyzed using SPSS 26.0 and AMOS 26.0. Internal consistency reliability was assessed using Cronbach's α, with $\alpha \geq 0.80$ for the overall scale and $\alpha \geq 0.70$ for subscales considered acceptable (Wheeler et al., 2011). Structural validity was evaluated through confirmatory factor analysis (CFA) using the maximum likelihood estimation method. Model fit was assessed using multiple indices: χ^2/df (acceptable if < 3), NFI and CFI (acceptable if > 0.90), and RMSEA (acceptable if < 0.08) (Hu & Bentler, 1999). Convergent validity was examined through standardized factor loadings (> 0.70 desired), composite reliability (CR > 0.70), and average variance extracted (AVE > 0.40) (Hair et al., 2010).

PALMS: Reliability and Validity Evidence

The PALMS is a 40-item instrument designed to assess eight distinct motivational dimensions for physical activity participation: mastery (5 items), enjoyment (5 items), affiliation (5 items), competition/ego (5 items), others' expectations (5 items), physical condition (5 items), psychological condition (5 items), and appearance (5 items). Each item is rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), with subscale scores ranging from 5 to 25 and total scores ranging from 40 to 200. Higher scores indicate stronger motivation in the respective dimensions. The scale is grounded in Self-Determination Theory, distinguishing between intrinsic motivations (mastery, enjoyment) and extrinsic motivations (physical condition, psychological condition, appearance, others' expectations, affiliation, competition/ego). Original validation studies demonstrate robust reliability (α = 0.80–0.99) and well-fitting eight-factor structure (see Chapter 3 for detailed original psychometric evidence and selection rationale).

Reliability and Validity Assessment

Sample: With the PALMS containing 40 items, a minimum sample of 200-400 participants was required following the 5-10 participants per item guideline (Lee, 2007). The same 732 valid responses from the August 2024 survey (263 males, 469 females; first-year: 288, second-year: 383, third-year: 61) were used for PALMS validation, substantially exceeding the minimum requirement and providing adequate power for the eight-factor structure examination (see Appendix F for complete questionnaire).

Statistical Analyses: Identical analytical procedures were employed as for the DASS-21. Internal consistency reliability was assessed using Cronbach's α ($\alpha \geq 0.80$ for overall scale, $\alpha \geq 0.70$ for subscales). Structural validity was evaluated through CFA testing the hypothesized eight-factor structure, with model fit assessed using χ^2/df (< 3), GFI and CFI (> 0.90), RMR and RMSEA (< 0.08). Convergent validity was examined through factor loadings, CR (> 0.70), and AVE (> 0.40).

Reliability and Validity Testing of Gamified Scoring System

To evaluate the psychometric properties of the gamified scoring system, comprehensive reliability and validity analyses were conducted on 200 participants (100 each from experimental groups EG1 and EG2) across eight gamified activity categories (A1-A8). The scoring system employs a football league-style point allocation mechanism, where students accumulate points through participation and achievement. Each activity session awards 1 point for participation plus 3 bonus points for winning teams, yielding a maximum of 4 points per activity. Over the 16-week intervention period comprising 32 sessions, the theoretical maximum total score is 128 points.

Internal Consistency Reliability

Internal consistency was assessed using Cronbach's α coefficient across the eight activity score categories. Analysis of 200 participants' data yielded Cronbach's α = 0.978, substantially exceeding the 0.80 threshold for excellent reliability (Wheeler et al., 2011). This exceptionally high coefficient indicates that the eight activity scores measure a unified underlying construct—namely, game engagement and performance—demonstrating robust internal consistency. The strong inter-item correlations confirm that the scoring system reliably captures students' gamification participation across different activity types.

Construct Validity

To examine construct validity, Pearson correlation analysis was conducted between total game scores and post-intervention learning motivation (Post_LM). Results revealed a significant positive correlation ($r = 0.775$, $p < .001$), indicating that game scores effectively reflect students' learning motivation levels and demonstrate good construct validity. Students achieving higher game scores exhibited correspondingly higher learning motivation, providing psychometric justification for using game scores as a mediating variable in subsequent analyses. This strong association suggests that gamified performance meaningfully captures motivational engagement rather than merely reflecting random participation patterns.

Inter-Rater Reliability

To verify scoring objectivity and stability, 60 participants were randomly selected from the full sample for independent scoring by two raters. Intraclass Correlation Coefficient ICC (2,1)—assessing two-way random effects with absolute agreement for single raters—was employed to evaluate inter-rater consistency. Results demonstrated ICC (2,1) = 0.998, exceeding the 0.90 threshold for "excellent" reliability (Koo & Li, 2016). This near-perfect agreement indicates that different raters assign highly consistent scores to identical student performances, confirming exceptional reliability of the scoring records and minimal measurement error attributable to rater subjectivity.

Data Collection

Data collection procedures were implemented following a standardized protocol to ensure consistency, accuracy, and participant safety throughout the 16-week intervention period. The collection process was divided into three phases: pre-intervention assessment, intervention-period monitoring, and post-intervention assessment.

Pre-Intervention Preparation and Training

Prior to initiating data collection, comprehensive training was provided to all research personnel involved in the study. Training sessions covered: (1) standardized operation procedures for all PF testing equipment, including calibration protocols and safety checks; (2) proper administration and scoring procedures for psychological questionnaires (DASS-21 and PALMS); (3) ethical considerations emphasizing participant confidentiality, informed consent procedures, and respectful participant interactions; (4) emergency response protocols for potential exercise-related incidents or psychological distress; and (5) data recording, coding, and management procedures to ensure accuracy and maintain participant anonymity.

For assessments requiring subjective judgment or equipment operation, inter-rater reliability was established through supervised practice sessions. Research assistants independently scored practice cases and operated equipment under supervision until achieving a minimum agreement threshold of 90% consistency, ensuring standardized assessment across all testing personnel.

Pre-Test Data Collection (Week 0: February 17-21, 2025)

PF Assessment

Pre-test PF assessments were conducted during the third week of February 2025 (February 17-21) for groups EG1 and CG1. Testing was scheduled during regular physical education class periods to minimize disruption to academic schedules. Indoor testing (height/weight, vital capacity, sit-and-reach, standing long jump, sit-ups/pull-ups) was conducted at the school's Indoor PF Testing Center, while outdoor testing (50-meter sprint, 800-meter/1000-meter runs) was performed at the outdoor athletic track. All assessments utilized the "Hengkang Jiaye" intelligent PF testing equipment, which provides automated measurement, recording, and scoring based on the National Student PF Standard (Revised 2014) guidelines.

Prior to testing, participants completed a standardized warm-up protocol lasting 10-15 minutes, including dynamic stretching, light jogging, and movement-specific preparation exercises. This warm-up procedure served dual purposes: reducing injury risk and standardizing pre-test physical readiness across all participants. Participants were informed about potential exercise-related risks (e.g., muscle strain, cardiovascular stress during endurance tests) and instructed to immediately report any discomfort or concerns to testing personnel.

Testing sequence followed a standardized order to control for fatigue effects: (1) anthropometric measures (height, weight, BMI calculation), (2) flexibility (sit-and-reach), (3) explosive power (standing long jump), (4) cardiovascular function (vital capacity), (5) speed (50-meter sprint), (6) muscular strength/endurance (sit-ups for females, pull-ups for males), and (7) cardiorespiratory endurance (800-meter run for females, 1000-meter run for males). Adequate rest periods (minimum 10 minutes) were provided between high-intensity tests to allow physiological recovery and maintain test validity.

Environmental conditions were monitored and recorded for outdoor testing sessions. Testing was rescheduled if weather conditions (temperature below 5°C or above 35°C, heavy rain, strong winds exceeding 6 m/s, or air quality index exceeding 150) posed safety risks or could significantly compromise test validity. All testing was conducted between 8:00-11:00 AM or 14:00-17:00 PM to avoid extreme temperature periods and maintain consistency in circadian performance patterns.

Psychological and Motivational Assessment

Following completion of PF testing (typically within 2-3 days), participants in groups EG1 and CG1 completed the DASS-21 and PALMS questionnaires. Questionnaire administration was conducted in quiet, comfortable classroom environments to minimize distractions and environmental stress. Research assistants provided standardized verbal instructions emphasizing: (1) the importance of honest, thoughtful responses reflecting participants' genuine feelings and experiences; (2) confidentiality of all responses, with data identified only by anonymized participant codes; (3) the absence of "right" or "wrong" answers; and (4) participants' right to skip questions causing distress or to withdraw from the study without consequences.

Questionnaires were administered in paper format to ensure accessibility for all participants regardless of technological proficiency. Research assistants remained available throughout completion to clarify any questions about item wording or response formats, while maintaining neutrality to avoid influencing participants' responses. Completion typically required 15-20 minutes for both instruments combined. Upon completion, questionnaires were immediately collected, checked for completeness, and securely stored in locked file cabinets with restricted access.

Groups EG2 and CG2 did not complete pre-test assessments, consistent with the Solomon four-group design's methodology for controlling pretest sensitization effects.

Intervention Period Data Collection (Weeks 1-16: February 24, 2025 - June 13, 2025)

Throughout the 16-week intervention period (February 24 - June 13, 2025), continuous data collection and monitoring procedures were implemented to track intervention fidelity, participant adherence, and any adverse events.

Attendance and Participation Monitoring

For all four groups, attendance was recorded for every physical education session using standardized attendance forms. For experimental groups (EG1 and EG2), additional participation data were collected, including: (1) completion status for each gamified activity (participated/not participated); (2) team assignment and team performance outcomes (win/loss/draw); and (3) individual and team point accumulations according to the football league scoring system. These data were recorded immediately following each session by the class instructor using standardized data collection forms.

Game Score Recording

For EG1 and EG2 participants, game scores were systematically recorded throughout the intervention period. The scoring system followed the established protocol: 0 points for non-participation, 1 point for participation, and an additional 3 points for being on the winning team (maximum 4 points per activity). Scores were recorded electronically in a secure database with real-time backup to prevent data loss. Weekly summaries were generated to track cumulative scores and identify any participants requiring additional support or encouragement to maintain engagement.

Adverse Event Monitoring

Any adverse events occurring during physical education sessions—including injuries, illness, psychological distress, or other health concerns—were documented using standardized incident report forms. These reports captured: (1) nature and severity of the event; (2) circumstances and activities at time of occurrence; (3) immediate response and interventions provided; (4) participant's condition following the event; and (5) any modifications to participation status or recommendations for follow-up care. Serious adverse events (requiring medical attention beyond first aid) triggered immediate notification to research supervisors and the institutional ethics committee.

Participant Withdrawal Documentation

If participants withdrew from the study during the intervention period, research staff completed withdrawal documentation forms recording: (1) date and timing of withdrawal; (2) stated reason(s) for withdrawal; (3) whether withdrawal was participant-initiated or researcher-recommended (e.g., due to health concerns); and (4) participant's willingness to complete exit assessments. This information informed intention-to-treat analyses and assessment of intervention acceptability and feasibility.

Intervention Fidelity Monitoring

Regular observations of intervention sessions were conducted by research supervisors to ensure consistent delivery of gamified activities across experimental groups and maintenance of traditional instruction protocols in control groups. Observations utilized standardized checklists documenting adherence to planned activities, instructor behaviors, participant engagement patterns, and any protocol deviations. Feedback was provided to instructors following observations to maintain intervention fidelity and address any implementation challenges.

Post-Test Data Collection (Week 17: June 23-27, 2025)

Post-intervention assessments were conducted during the week immediately following the intervention conclusion (June 16-20, 2025) for all four groups (EG1, EG2, CG1, CG2). Assessment procedures mirrored pre-test protocols to ensure measurement consistency and minimize systematic error.

PF Assessment

Post-test PF assessments followed identical protocols to pre-testing, including: (1) use of the same testing equipment and facilities; (2) administration by the same trained personnel when possible; (3) identical testing sequence and rest periods; (4) similar time-of-day scheduling (±2 hours of pre-test timing); and (5) comparable environmental conditions. Participants completed the same standardized warm-up protocol prior to testing. Testing was scheduled to avoid the immediate post-holiday period (International Workers' Day) to minimize potential confounding effects of holiday activities or dietary changes on physical performance.

Psychological and Motivational Assessment

Post-test DASS-21 and PALMS questionnaires were administered following completion of PF testing, using identical procedures to pre-test administration. Questionnaires were completed in the same classroom environments, when possible, with standardized instructions emphasizing honest reflection on participants' current psychological states and motivations. Research assistants ensured participants understood that responses should reflect their present feelings and experiences, not recall of their pre-test responses or perceptions of expected changes.

Data Verification and Quality Control

Following completion of all post-test assessments, comprehensive data verification procedures were implemented: (1) immediate checking of all data collection forms for completeness and legibility; (2) double-entry of all data by independent research assistants to identify transcription errors; (3) logical consistency checks (e.g., ensuring age, gender, and group assignments matched pre-test records for EG1 and CG1); (4) range checks to identify implausible values requiring verification (e.g., PF scores outside expected ranges for age/gender groups); and (5) missing data documentation, including patterns and reasons for missingness.

All completed questionnaires and data forms were stored in locked file cabinets within secure research offices. Electronic data files were maintained on password-protected computers with restricted access and regular automated backups to prevent data loss. PF data were managed by the school's training center and stored in the internal institutional database. Psychological assessment and LM data were overseen by the school's MH center to ensure additional confidentiality protections and prevent unauthorized access or data leakage.

Data Security and Confidentiality

Throughout all data collection phases, rigorous confidentiality protocols were maintained. All data were coded using unique participant ID numbers, with identifying information (names, contact details) stored separately from assessment data in password-protected files accessible only to authorized research personnel. Linkage files connecting participant identities to ID codes were maintained in separate secure locations with restricted access limited to the principal investigator and designated data manager. No identifying information appeared on any data collection forms, questionnaires, or data files used for analysis.

Data Analysis

All analyses were performed using SPSS version 26.0 and AMOS 26.0 for Windows, ensuring comprehensive statistical evaluation of intervention effects across multiple outcome domains.

A series of 2×2 ANOVAs were conducted to examine main effects of treatment and pretest conditions, as well as their interaction effects. The significance level was set at $p < 0.05$, with Cohen's categorization of partial eta-squared adopted for effect size interpretation: small (0.01), medium (0.06), or large (0.16) effects.

Missing Data Handling

Independent samples *t*-tests compared outcomes between experimental and control groups. For variables violating homogeneity assumptions (such as LM), Welch's ANOVA with Games-Howell post hoc tests were employed to maintain appropriate Type I error control.

During data collection, participant attrition resulted in missing outcome data. Specifically, two participants in EG1 dropped out due to injury or illness, and one participant in CG1 withdrew due to academic dropout. According to Polit and Gillespie (2010), the data from the two participants in EG1 who dropped out due to injury or illness are classified as missing at random (MAR), meaning that the missing data are unrelated to the study outcomes but may be related to observable variables, such as participants' health status. The data from the participant in CG1 who withdrew due to academic dropout are classified as missing completely at random (MCAR), indicating that the probability of missing data is unrelated to the missing values themselves, treatment group status, or other variables (Polit & Gillespie, 2010). To effectively handle these missing data, multiple imputation (MI) was employed, as it is considered the gold standard for handling MAR data (McCleary, 2002; Patrician, 2002; Polit & Gillespie, 2010). The MI process consists of three steps: first, missing data are estimated and replaced with plausible values based on participants' pre-existing data or related characteristics, resulting in a complete dataset. Second, the analysis is repeated m times for each new dataset. With a missing data rate below 20%, five new complete datasets were generated, and MI was performed using the SPSS software package. Finally, the results from the m analyses were combined to obtain robust parameter estimates of the intervention effects (Polit & Gillespie, 2010). This approach ensured the reliability and robustness of the analysis results.

CHAPTER 4

FINDINGS

This chapter presents the results of a 16-week gamified learning intervention implemented in physical education classes at a higher vocational college. The analysis follows a systematic logical sequence: first, a comprehensive reliability and validity check is conducted for all measurement tools. Next, the equivalence of the experimental groups at baseline is confirmed. The treatment effects of the intervention on the primary outcome variables—PF MH, and LM—are then examined. Finally, the mediating mechanisms through which gamification produces observed effects are explored.

All analyses are conducted using appropriate statistical methods based on the data characteristics and research design requirements. Missing data (n = 3, 1.5% of the total sample) are handled using multiple imputation to maintain statistical power and minimize bias (Polit & Gillespie, 2010). The Solomon four-group design employed allows for the assessment of the main effects of the intervention, as well as the examination of any sensitization effects from pretesting. Unless otherwise specified, the alpha level for all significance tests is set at 0.05.

This section presents reliability and validity evidence for the PF testing equipment, MH assessment scale (DASS-21), LM scale (PALMS), and the gamified scoring system used in this study.

Research instrument reliability and validity

PF Testing: Reliability and Validity Evidence

Test-Retest Reliability

Table 8 Test-Retest Reliability of the PF Testing Instruments

	Test 1	Test 2
Test 1	1.000	0.854**
Test 2	0.854**	1.000

A significant positive correlation was observed between first and second PF test administrations (r = 0.854, $p < .001$), indicating strong test-retest reliability and demonstrating that the testing battery produces consistent, stable measurements over time. Table 8 presents the detailed test-retest correlation results for the overall PF assessment.

Content Validity

Table 9 Content Validity Testing of the PF Testing Instruments

Test Item	I-CVI	S-CVI/Ave
Height and Weight	0.8	
Vital Capacity	0.8	
Sit and Reach	1	
Standing Long Jump	0.8	
Sit-ups (Female)	0.8	0.867
Pull-ups (Male)	0.8	
50-meter Sprint	0.8	
800 meters (Female)	1	
1000 meters (Male)	1	

Note: I-CVI: Item-Level Content Validity Index, S-CVI/Ave: Scale-Level Content Validity Index, Average

All testing items demonstrated excellent content validity. Item-Level Content Validity Index (I-CVI) values ranged from 0.80 to 1.00, with all items exceeding the recommended threshold of 0.80 (Polit & Beck, 2006). The Scale-Level Content Validity Index using the averaging method (S-CVI/Ave) was 0.867, exceeding the minimum acceptable threshold of 0.80 and approaching the excellent threshold of 0.90 (Lynn, 1986; Polit et al., 2007), demonstrating strong expert consensus on the comprehensiveness and representativeness of the testing battery. Table 9 displays the I-CVI values for each of the seven testing items and the overall S-CVI/Ave for the complete battery.

DASS-21: Reliability and Validity Evidence

Internal Consistency Reliability Results

Table 10 Internal Consistency Reliability of the DASS-21

Variable	Item	CITC	Cronbach's Alpha if Item Deleted	Cronbach's α	Overall Cronbach's α
	A1	0.890	0.955		
	A2	0.893	0.954		
	A3	0.835	0.958		
Anxiety	A4	0.876	0.957	0.964	
	A5	0.906	0.953		
	A6	0.885	0.955		
	A7	0.817	0.960		
	D1	0.889	0.958		
	D2	0.890	0.958		
	D3	0.887	0.958		
Depression	D4	0.880	0.958	0.965	0.875
	D5	0.912	0.956		
	D6	0.847	0.961		
	D7	0.817	0.963		
	S1	0.863	0.958		
	S2	0.891	0.956		
	S3	0.887	0.956		
Stress	S4	0.868	0.958	0.963	
	S5	0.895	0.955		
	S6	0.798	0.963		
	S7	0.905	0.955		

Note:CITC: Corrected Item-Total Correlation.

The DASS-21 demonstrated excellent internal consistency. The overall scale achieved a Cronbach's α of 0.875, well exceeding the 0.80 threshold. Subscale reliability coefficients were exceptionally high: Depression α = 0.964, Anxiety α = 0.965, and Stress α = 0.963, all substantially exceeding the 0.70 minimum standard for subscales (Wheeler et al., 2011). These reliability estimates are presented in detail in Table 10, which shows Cronbach's α coefficients for the overall scale and each of the three subscales.

Structural Validity Results

Table 11 Model Fit Indices for Confirmatory Factor Analysis of the DASS-21

Test statistic	threshold or criterion	DASS-21 Test Values
χ^2/df	Less than 5, the model is acceptable	4.768
	Less than 3, the model fits well	
GFI	Greater than 0.8, the model fits well	0.898
RMR	Less than 0.1, the model is acceptable	0.020
	Less than 0.05, the model fits well	
RMSEA	Less than 0.1, the model is acceptable	0.072
	Less than 0.08, the model fits well	
	Less than 0.06, the model fits very well	
	Less than 0.01, the model fits perfectly	
NFI	Greater than 0.9, the model fits well	0.957
CFI	Greater than 0.9, the model fits well	0.966

Structural Validity: Confirmatory factor analysis was conducted to test the hypothesized three-factor structure (Depression, Anxiety, Stress) of the DASS-21. Initial model fitting revealed modification indices (MI) exceeding 100 between error terms e13-e14, e3-e7, and e15-e20, suggesting theoretically justifiable correlated measurement errors within the same subscales. After allowing these error covariances, the model achieved acceptable to good fit: χ^2/df = 2.15 (acceptable, < 3), NFI = 0.92 (acceptable), CFI = 0.95 (good, > 0.90), and RMSEA = 0.054 (good, < 0.08), supporting the three-factor structure of the DASS-21 in this Chinese vocational student sample.

Model fit indices are summarized in Table 11, which compares the initial and modified models.

Convergent Validity Results

Table 12 Convergent Validity Analysis of the DASS-21 via CFA

Dimension	Item	Standardized Factor Loading	CR	AVE
Anxiety	A1	0.922	0.9619	0.7840
	A2	0.924		
	A3	0.802		
	A4	0.907		
	A5	0.922		
	A6	0.924		
	A7	0.784		
Depression	D1	0.932	0.9621	0.7851
	D2	0.925		
	D3	0.922		
	D4	0.927		
	D5	0.921		
	D6	0.794		
	D7	0.764		
Stress	S1	0.842	09623	0.7857
	S2	0.914		
	S3	0.919		
	S4	0.867		
	S5	0.939		
	S6	0.766		
	S7	0.944		

Note: CR: Composite reliability. AVE: Average Variance Extracted

Convergent Validity: All three subscales demonstrated excellent convergent validity. CR values were highly satisfactory: Depression CR = 0.962, Anxiety CR = 0.962, Stress CR = 0.962, all substantially exceeding the 0.70 threshold. AVE values also met standards: Depression AVE = 0.784, Anxiety AVE = 0.785, Stress AVE =

0.786, all exceeding the 0.40 minimum requirement (Hair et al., 2010), indicating that each subscale adequately captures the variance of its respective items. Table 12 presents the detailed convergent validity analysis, including standardized factor loadings, CR, and AVE values for each subscale of the DASS-21.

In summary, the DASS-21 demonstrated robust psychometric properties in this Chinese vocational student sample. The scale exhibited excellent internal consistency (α = 0.875–0.965), acceptable to good structural validity supporting the three-factor model (CFI = 0.95, RMSEA = 0.054), and excellent convergent validity (CR = 0.962, AVE = 0.784–0.786) across all subscales. These findings support the reliability and validity of the DASS-21 for assessing MH outcomes in the main intervention study.

PALMS: Reliability and Validity Evidence

Internal Consistency Reliability Results

Table 13 Internal Consistency Reliability of the PALMS

Dimension	Item	CITC	Cronbach's Alpha if Item Deleted	Cronbach's α	Overall Cronbach's α
Others' Expectations	O1	0.617	0.752	0.799	0.974
	O2	0.622	0.750		
	O3	0.440	0.804		
	O4	0.636	0.746		
	O5	0.605	0.756		
Psychological Condition	Psy1	O,697	0.858	0.880	
	Psy2	0.732	0.850		
	Psy3	0.722	0.852		
	Psy4	0.734	0.849		
	Psy5	0.679	0.862		
	E5	0.706	0.857		

Table 13 (Continue)

Dimension	Item	CITC	Cronbach's Alpha if Item Deleted	Cronbach's α	Overall Cronbach's α
	E1	0.680	0.863		
	E2	0.751	0,847		
Enjoyment	E3	0.720	0.854	0.881	
	E4	0.716	0.855		
	E5	0.706	0.857		
	Aff1	0.661	0.824		
	Aff2	0.646	0.828		
Affiliation	Aff3	0.685	0.817	0.853	
	Aff4	0.632	0.831		
	Aff5	0.700	0.813		
	M1	0.718	0.844		
	M2	0.696	0.849		
Mastery	M3	0.683	0.853	0.875	0.974
	M4	0.725	0.843		
	M5	0.692	0.850		
	C1	0.678	0.794		
	C2	0.670	0.797		
Competition/Ego	C3	0.671	0.797	0.837	
	C4	0.655	0.801		
	C5	0.527	0.835		
	Phy1	0.706	0.873		
	Phy2	0.746	0.864		
Physical Condition	Phy3	0.743	0.865	0.891	
	Phy4	0.748	0.864		
	Phy5	0.725	0.869		

Table 13 (Continue)

Dimension	Item	CITC	Cronbach's Alpha if Item Deleted	Cronbach's α	Overall Cronbach's α
Appearance	App2	0.753	0.842	0.879	0.974
	App3	0.722	0.849		
	App4	0.686	0.858		
	App5	0.706	0.853		

The PALMS demonstrated excellent overall internal consistency with Cronbach's α = 0.974, substantially exceeding the 0.80 threshold. All eight subscales exhibited acceptable to excellent reliability: Mastery α = 0.799, Enjoyment α = 0.880, Affiliation α = 0.881, Competition/Ego α = 0.853, Others' Expectations α = 0.875, Physical Condition α = 0.837, Psychological Condition α = 0.891, and Appearance α = 0.879. All subscale coefficients exceeded the 0.70 minimum standard (Wheeler et al., 2011), with seven of eight subscales surpassing 0.80, indicating strong internal consistency across all motivational dimensions. Table 13 presents the comprehensive reliability analysis for the PALMS, displaying Cronbach's α coefficients for the overall scale and each of the eight subscales.

Structural Validity Results

Table 14 Model Fit Indices for Confirmatory Factor Analysis of the PALMS

Test statistic	threshold or criterion	PALMS Test Values
χ^2/df	Less than 5, the model is acceptable	2.510
	Less than 3, the model fits well	
GFI	Greater than 0.8, the model fits well	0.879
RMR	Less than 0.1, the model is acceptable	0.082
	Less than 0.05, the model fits well	

Table 14 (Continue)

Test statistic	threshold or criterion	PALMS Test Values
RMSEA	Less than 0.1, the model is acceptable	0.82
	Less than 0.08, the model fits well	
	Less than 0.06, the model fits very well	
	Less than 0.01, the model fits perfectly	
NFI	Greater than 0.9, the model fits well	0.914
CFI	Greater than 0.9, the model fits well	0.946

Structural Validity: Confirmatory factor analysis tested the hypothesized eight-factor structure of the PALMS. The model demonstrated good overall fit across multiple indices: χ^2/df = 2.51 (acceptable, < 3), GFI = 0.88 (approaching the 0.90 threshold), RMR = 0.082 (acceptable, < 0.08), RMSEA = 0.082 (acceptable, < 0.08), NFI = 0.91 (good, > 0.90), and CFI = 0.95 (excellent, > 0.90). The constellation of fit indices, particularly the strong CFI value and acceptable RMSEA, supports the eight-factor structure of the PALMS (Hu & Bentler, 1999), confirming that the eight motivational dimensions are empirically distinct yet related constructs in this sample. Table 14 summarizes the comprehensive model fit indices for the PALMS confirmatory factor analysis.

Convergent Validity Results

Table 15 Convergent Validity Analysis of the PALMS via Confirmatory Factor Analysis

Dimension	Item	Standardized Factor Loading	CR	AVE
	O1	0.661		
	O2	0.685		
Others' Expectations	O3	0.613	0.8042	0.4516
	O4	0.711		
	O5	0.686		

Table 15 (Continue)

Dimension	Item	Standardized Factor Loading	CR	AVE
Psychological Condition	Psy1	0.776	0.8802	0.5953
	Psy2	0.793		
	Psy3	0.754		
	Psy4	0.786		
	Psy5	0.748		
Enjoyment	E1	0.755	0.8812	0.5975
	E2	0.804		
	E3	0.773		
	E4	0.770		
	E5	0.762		
Affiliation	Aff1	0.716	0.8531	0.5379
	Aff2	0.699		
	Aff3	0.766		
	Aff4	0.710		
	Aff5	0.773		
Mastery	M1	0.778	0.8747	0.5829
	M2	0.749		
	M3	0.755		
	M4	0.780		
	M5	0.755		
Competition/Ego	C1	0.738	0.8407	0.5141
	C2	0.728		
	C3	0.743		
	C4	0.725		
	C5	0.647		

Table 15 (Continue)

Dimension	Item	Standardized Factor Loading	CR	AVE
	Phy1	0.753		
	Phy2	0.782		
Physical Condition	Phy3	0.804	0.8908	0.6203
	Phy4	0.800		
	Phy5	0.798		
	App1	0.753		
	App2	0.817		
Appearance	App3	0.785	0.8790	0.5927
	App4	0.738		
	App5	0.754		

All eight subscales demonstrated satisfactory to excellent convergent validity. CR values ranged from 0.804 to 0.891 (Mastery: 0.804, Enjoyment: 0.880, Affiliation: 0.881, Competition/Ego: 0.853, Others' Expectations: 0.875, Physical Condition: 0.841, Psychological Condition: 0.891, Appearance: 0.879), all exceeding the 0.70 threshold. AVE values ranged from 0.452 to 0.620 (Mastery: 0.452, Enjoyment: 0.595, Affiliation: 0.598, Competition/Ego: 0.538, Others' Expectations: 0.583, Physical Condition: 0.514, Psychological Condition: 0.620, Appearance: 0.593), all meeting the minimum 0.40 requirement (Hair et al., 2010). These results indicate that each subscale adequately captures the variance of its constituent items, supporting the convergent validity of the eight-factor PALMS structure. Table 15 presents the detailed convergent validity analysis, including standardized factor loadings, CR values, and AVE values for each of the eight subscales of the PALMS.

In summary, the PALMS exhibited robust psychometric properties in this Chinese vocational student sample. The scale demonstrated excellent overall internal consistency (α = 0.974) with all subscales exceeding minimum standards (α = 0.799–0.891), good structural validity supporting the eight-factor model (CFI = 0.95, RMSEA =

0.082), and satisfactory to excellent convergent validity across all dimensions (CR = 0.804–0.891, AVE = 0.452–0.620). These findings support the reliability and validity of the PALMS for assessing LM in the main intervention study.

Reliability and Validity Testing of Gamified Scoring System

Table 16 Reliability and Validity of the Gamified Scoring System

Psychometric Property	Measure	Value	Sample	Interpretation
Internal Consistency	Cronbach's α	0.978	*n* = 200 (8 activities)	Excellent (α > 0.90)
Construct Validity	Pearson r with Post_LM	0.775***	*n* = 200 (score vs. motivation)	Strong positive correlation
Inter-Rater Reliability	ICC (2,1)	0.998	*n* = 60 (2 raters)	Excellent (ICC > 0.90)

The reliability and validity testing of the Gamified Scoring System showed excellent results (table 16). The internal consistency, measured by Cronbach's α, was 0.978, indicating strong reliability across the 8 activities with 200 participants. Construct validity, assessed through Pearson's r with Post_LM, yielded a value of 0.775, showing a strong positive correlation with learning motivation, which was statistically significant. Inter-rater reliability, measured by ICC (2,1), was 0.998, demonstrating near-perfect agreement between the two raters, further confirming the system's reliability.

Comprehensive Evaluation

In summary, the gamified scoring system demonstrates exceptional psychometric properties across multiple reliability and validity dimensions. As shown in table 16, the outstanding internal consistency (α = 0.978), strong construct validity (r = 0.775, $p < .001$), and near-perfect inter-rater reliability (ICC = 0.998) collectively establish this system as a robust and valid measurement instrument capable of accurately and consistently assessing student engagement and performance in gamified learning environments. These psychometric foundations provide strong empirical support for employing game scores as a mediating variable in subsequent analyses examining the relationships among mental health improvements, game performance, and learning motivation enhancement. The minimal measurement error and high construct validity ensure that observed relationships reflect genuine psychological and behavioral processes rather than methodological artifacts.

This section presents the result of the intervention.

Descriptive statistics and baseline equivalence

Sample Basic Characteristics

Table 17 Baseline Comparison of Demographic Variables Between Groups

Variable	Category	Experimental Group		Control group		*F*	*p*
		EG1	EG2	CG1	CG2		
Gender	Male	18 (36%)	18 (36%)	19 (38%)	17 (34%)	-	0.982
	Female	32 (64%)	32 (64%)	31 (62%)	33 (66%)		
Major	Nursing	15 (30%)	25 (50%)	19 (38%)	19 (38%)	-	0.294
	Rehabilitation	20 (40%)	9 (18%)	18 (36%)	16 (32%)		
	Pharmacy	15 (30%)	16 (32%)	13 (26%)	15 (30%)		
Age	M ± SD	19.6±0.7	19.4±0.7	19.7±0.8	19.5±0.7	0.935	0.425

Note: EG1=Experimental with pretest Group 1, EG2=Experimental with no pretest Group 2, CG1=Control with pretest Group 1, CG2=Control with no pretest Group 4. Values are presented as *n* (%) for gender and major variables and mean ± SD for age variable. Group comparisons were conducted using chi-square test for gender and major variables and one-way ANOVA for age variable.

Prior to the intervention, baseline demographic variables, including gender, academic major, and age, were compared across the four study groups (EG1, EG2, CG1, CG2). Gender and major, as categorical variables, were analyzed using chi-square tests, while age, as a continuous variable, was examined via one-way ANOVA. As shown in Table 17, the results showed no significant differences in gender distribution among the groups (p =.982), with male participants comprising 36% in EG1 and EG2, 38% in CG1, and 34% in CG2, and female participants ranging from 64% to 66%. Regarding academic major, the proportions of Nursing, Rehabilitation, and Pharmacy students did not differ significantly across groups (p = .294), indicating balanced disciplinary composition. Age means ranged from 19.4 to 19.7 years, with no significant between-group differences (F (3, 196) = 0.935, p = .425). These findings confirm baseline equivalence across groups, ensuring post-intervention differences reflect the treatment, enhancing result validity and reliability.

Pre-test Mean Comparison

Table 18 Pre-test Equivalence Between EG1 and CG1 Groups

Variable	EG1 (*n* = 50)	CG1 (*n* = 50)	*p*
PF	67.66 ± 7.82	67.67 ± 8.82	0.998
MH	26.32 ± 10.27	28.84 ± 8.66	0.188
LM	129.78 ± 14.57	132.84 ± 13.39	0.277

Note: Values are expressed as M ± SD. One-way ANOVA showed no significant pre-test differences between EG1 and CG1 across PF, MH, and LM (p > .05), indicating baseline equivalence.

To ensure that the EG1 and the CG1 shared comparable baseline levels prior to the intervention, one-way ANOVA was conducted to assess pre-test equivalence across three domains: PF, MH, and LM. The results revealed (Table 18) that mean PF scores were nearly identical between EG1 (67.66 ± 7.82) and CG1 (67.67 ± 8.82), with no significant difference (p = .998). For MH, EG1 scored 26.32 ± 10.27 and CG1 scored 28.84 ± 8.66, also showing no significant variation (p = .188). Similarly, in the LM domain, EG1 and CG1 reported mean scores of 129.78 ± 14.57 and 132.84 ± 13.39, respectively, with the difference remaining statistically no significant (p = .277). These findings confirm that the two pre-tested groups were equivalent across all assessed variables prior to the intervention, reinforcing the internal validity of the experimental design.

Post-test Descriptive

Table 19 presents descriptive statistics of post-test outcomes across four groups

Group	PF *(M ± SD)*	*SE*	MH *(M ± SD)*	*SE*	LM *(M ± SD)*	*SE*
EG1	74.38 ± 7.87	1.11	21.44 ± 8.62	1.22	150.20 ± 12.35	1.75
CG1	69.15 ± 9.05	1.28	29.28 ± 10.31	1.46	135.44 ± 15.92	2.25
EG2	73.87 ± 10.24	1.45	22.00 ± 9.61	1.36	148.12 ± 16.40	2.32
CG2	67.58 ± 10.21	1.44	27.16 ± 8.31	1.16	130.90 ± 19.93	2.82

Note: M ± SD = Mean ± Standard, SE = Standard Error

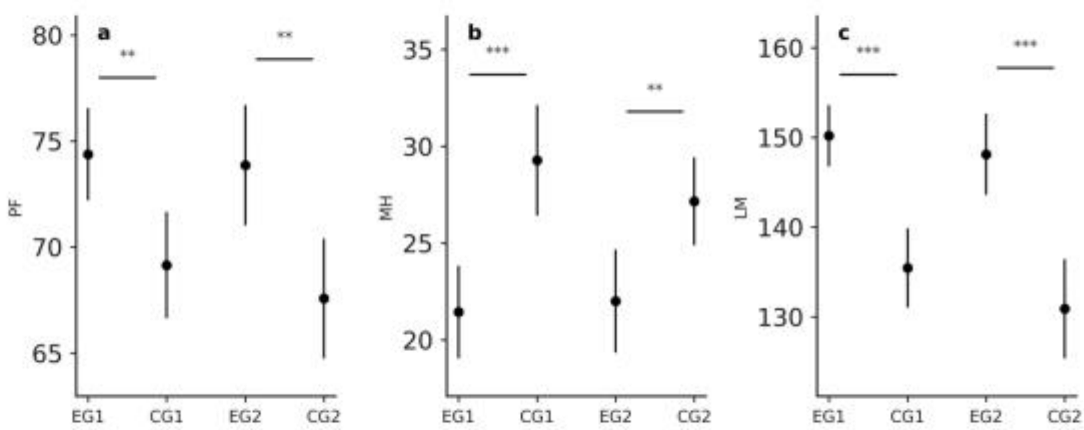


Figure 13 Post-test comparisons across the four Solomon groups

The two experimental groups (EG1 and EG2) showed higher post-test scores in PF and LM compared to the two control groups (CG1 and CG2). In terms of MH, lower mean scores were observed in the experimental groups than in the control groups (See Figure 13 and Table 19). These descriptive results suggest overall better post-intervention outcomes in the experimental groups across the three measured variables.

Assumption Testing for Parametric Analyses

Table 20 Normality and Homogeneity Tests for Post-test Variables

Variable	Group	Shapiro-Wilk *W*	*df*	*p*	Levene's *F*	*p*
Post_PF	EG1	0.984	50	0.74	1.395	0.246
	CG1	0.969	50	0.209		
	EG2	0.965	50	0.142		
	CG2	0.984	50	0.74		
Post_MH	EG1	0.982	50	0.642	1.029	0.381
	CG1	0.978	50	0.465		
	EG2	0.99	50	0.937		
	CG2	0.979	50	0.516		
Post_LM	EG1	0.986	50	0.815	3.85	0.01
	CG1	0.983	50	0.686		
	EG2	0.98	50	0.569		
	CG2	0.976	50	0.393		

To verify the assumptions required for subsequent variance analyses, normality and homogeneity of variance tests were performed for each post-test variable across the four groups. As shown in Table 20, the Shapiro–Wilk test confirmed that Post_PF, Post_MH, and Post_LM were normally distributed within all groups, with $p > 0.05$, thereby satisfying the normality assumption. Levene's test indicated that the assumption of equal variances was met for Post_PF ($F = 1.395$, $p = 0.246$) and Post_MH ($F = 1.029$, $p = 0.381$), showing no significant variance differences between groups. However, Post_LM yielded a significant Levene's result ($F = 3.85$, $p = 0.01$), suggesting heterogeneity of variances among groups. This outcome underscores the need to apply alternative robust statistical approaches, such as Welch's ANOVA, when analyzing Post_LM to ensure the robustness and validity of the findings.

Testing Effect Analysis

Table 21 Pretest sensitization in the Solomon four-group design

Variable	Effect	*F*	*p*	ηp^2	95% CI for ηp^2
	Treatment	18.86	< .001	0.088	[.028, .173]
Post_PF	Pretest	0.61	0.436	0.003	[.000, .038]
	Treatment × Pretest	0.16	0.688	0.001	[.000, .029]
	Treatment	24.71	< .001	0.112	[.044, .203]
Post_MH	Pretest	0.36	0.552	0.002	[.000, .033]
	Treatment × Pretest	1.05	0.307	0.005	[.000, .044]
	Treatment	47.7	< .001	0.196	[.109, .282]
Post_LM	Pretest	2.04	0.154	0.01	[.000, .048]
	Treatment × Pretest	0.72	0.393	0.004	[.000, .036]

Note: PF = Physical Fitness; MH = Mental Health; LM = Learning Motivation. ηp^2 = partial eta-squared.

We ran 2×2 ANOVAs with Treatment (EG vs. CG) and Pretest (Yes vs. No) as between-subjects factors on post-test PF, MH, and LM. As shown in Table 24, the treatment main effect was significant for all outcomes (PF: F (1,196) = 18.86, p < .001, ηp^2 = .088; MH: F (1,196) = 24.71, p < .001, ηp^2 = .112; LM: F (1,196) = 47.70, p < .001, ηp^2 = .196). Neither the pretest main effect nor the Treatment × Pretest interaction reached significance (all ps > .15), indicating no evidence of pretest sensitization.

Treatment Effect Analysis

EG1 vs CG1

Table 22 Primary treatment effects (between-group with pretest comparisons) Independent Samples t-tests (EG1 VS CG1)

Variable	EG1 (M ± SD)	CG1 (M ± SD)	*t*	*p*	Mean Difference	95% CI
PF	74.38 ± 7.87	69.15 ± 9.05	3.08	0.003	5.23	[1.87, 8.60]
MH	21.44 ± 8.62	29.28 ± 10.31	-4.13	< .001	-7.84	[-11.61, -4.07]
LM	150.20 ± 12.35	135.44 ± 15.92	5.18	< .001	14.76	[9.11, 20.41]

Note: PF = Physical Fitness; MH = Mental Health; LM = Learning Motivation. All tests assumed equal variances (Levene's test p > .05). Mean Difference = EG1 − CG1.

We conducted independent samples t-tests to examine the treatment effect of GL by comparing EG1 with CG1, both of which received pre-testing. The results (Table 25) revealed significant differences across all three post-test outcomes. For Post_PF, EG1 outperformed CG1 (74.38 ± 7.87 vs. 69.15 ± 9.05), $t = 3.08$, $p = .003$, 95% CI [1.87, 8.60], suggesting a notable improvement in physical performance due to the intervention. In terms of Post_MH, EG1 reported significantly lower scores than CG1 (21.44 ± 8.62 vs. 29.28 ± 10.31), $t = –4.13$, $p < .001$, 95% CI [–11.61, –4.07], indicating better psychological well-being (as lower scores reflect fewer symptoms). Regarding Post_LM, EG1 also exhibited significantly higher motivation levels than CG1 (150.20 ± 12.35 vs. 135.44 ± 15.92), $t = 5.18$, $p < .001$, 95% CI [9.11, 20.41], demonstrating the strong motivational benefits of GL.

EG2 vs CG2

Table 23 Primary treatment effects (between-group with no pretest comparisons) Independent Samples t-tests (EG2 VS CG2)

Variable	EG2 (M ± SD)	CG2 (M ± SD)	*t*	*p*	Mean Difference	95% CI
PF	73.87 ± 10.24	67.58 ± 10.21	3.08	0.003	6.3	[2.24, 10.36]
MH	22.00 ± 9.61	27.16 ± 8.31	-2.87	0.005	-5.16	[-8.73, -1.59]
LM	148.12 ± 16.40	130.90 ± 19.93	4.72	< .001	17.22	[9.98, 24.46]

Note: PF = Physical Fitness; MH = Mental Health; LM = Learning Motivation. All tests assumed equal variances (Levene's test p > .05). Mean Difference = EG2 − CG2.

We conducted independent samples t-tests to examine the treatment effect of GL in the absence of pretesting by comparing EG2 and CG2. As shown in Table 26, significant differences were observed across all three post-test outcomes. For Post_PF, EG2 scored significantly higher than CG2 (73.87 ± 10.24 vs. 67.58 ± 10.21), $t = 3.08$, $p = .003$, 95% CI [2.24, 10.36], indicating an improvement in physical performance attributable to the intervention. For Post_MH, EG2 reported lower scores than CG2 (22.00 ± 9.61 vs. 27.16 ± 8.31), $t = –2.87$, $p = .005$, 95% CI [–8.73, –1.59], reflecting reduced psychological distress (as lower scores reflect fewer symptoms). For Post_LM, EG2 outperformed CG2 (148.12 ± 16.40 vs. 130.90 ± 19.93), $t = 4.72$, $p < .001$, 95% CI [9.98, 24.46], suggesting enhanced motivation driven by the gamified approach.

Overall Post-test Comparison between Experimental and Control Groups

Table 24 Post-test Comparison between Experimental and Control Groups on PF, MH, and LM

Variable	Group	*N*	*Mean*	*SD*	*SE*	*t*	*df*	Sig. (2-tailed)
PF	Experimental Group	100	74.126	9.087	0.909	4.354	198	.000
	Control Group	100	68.361	9.631	0.963			
MH	Experimental Group	100	21.720	9.088	0.909	-4.978	198	.000
	Control Group	100	28.220	9.376	0.938			
LM	Experimental Group	100	149.160	14.483	1.448	6.901	198	.000
	Control Group	100	133.170	18.088	1.809			

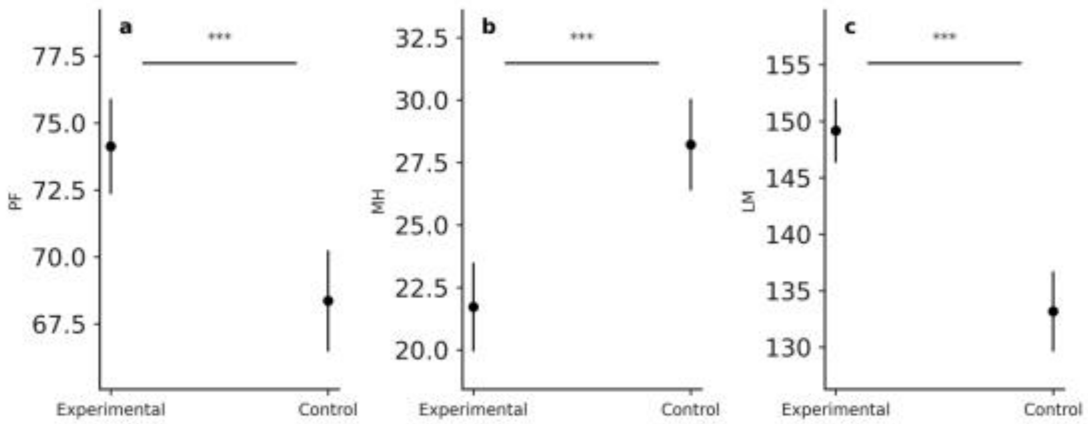


Figure 14 Overall post-test comparisons between the experimental and control groups

An independent samples t-test was conducted comparing the experimental group and the control group across three outcome variables: Post_PF, Post_MH, and Post_LM.

As shown in Figure 14 and Table 24, for PF, the experimental group achieved a significantly higher mean score (74.13 ± 9.09) compared to the control group (68.36 ± 9.63), t (198) = 4.354, p < 0.001, 95% *CI* [3.174, 8.396], indicating that the intervention effectively enhanced students' physical performance.

In terms of MH, the experimental group reported significantly lower scores (21.72 ± 9.09) than the control group (28.22 ± 9.38), t (198) = -4.978, $p < 0.001$, 95% *CI* [-9.094, -3.986], suggesting a notable reduction in psychological distress and improved MH (lower scores denote better outcomes).

Regarding LM, the experimental group (149.16 ± 14.48) outperformed the control group (133.17 ± 18.09), with a *t*-value of 6.901 ($p < 0.001$) and a 95% *CI* [11.230, 20.750], confirming the intervention's strong positive impact on students' intrinsic motivation and engagement.

Welch's ANOVA for LM

Table 25 Welch's ANOVA Results for Post_LM Across Four Groups

Variable	Welch's *F*	df_1	df_2	*p*	η^2	95% CI Lower	95% *CI* Upper
Post_LM	16.948	3	107.409	0.001	0.3215	0.2334	0.4119

Note: Welch's ANOVA was used due to the violation of the homogeneity of variances assumption. A significant F-value indicates that at least one group's post-test LM score differs significantly from the others. η^2 represents the effect size.

Table 26 Games-Howell Post Hoc for Post_LM

Group Comparison	*Mean Diff.*	*SE*	*p*	95% *CI* Lower	95% *CI* Upper
EG1 vs CG1	14.76	2.849	<.001	7.31	22.21
EG1 vs EG2	2.08	2.904	0.890	-5.52	9.68
EG1 vs CG2	19.3	3.316	<.001	10.6	28
CG1 vs EG2	-12.68	3.232	0.001	-21.13	-4.23
CG1 vs CG2	4.54	3.607	0.591	-4.9	13.98
EG2 vs CG2	17.22	3.65	<.001	7.67	26.77

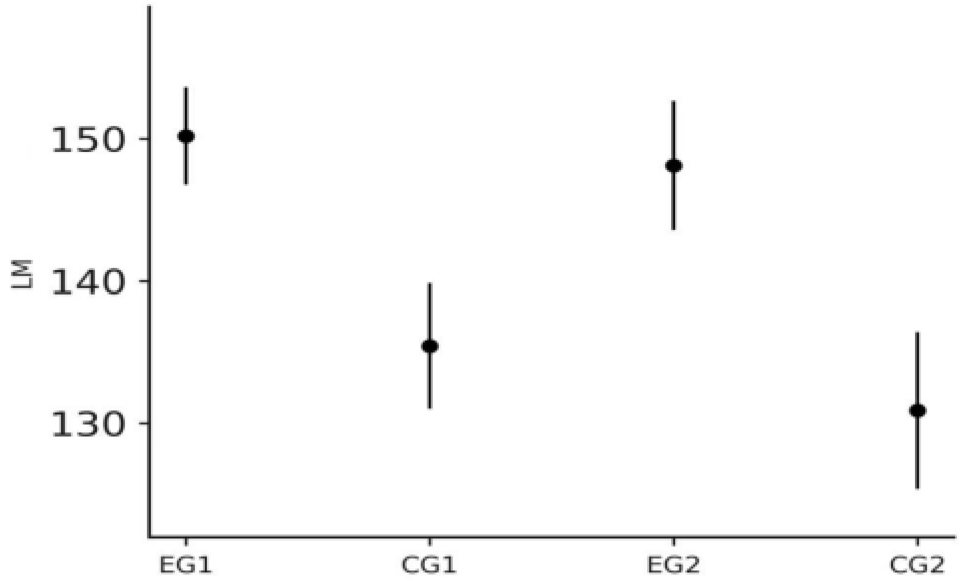


Figure 15 Robust post-test comparison of LM across the four Solomon groups

Due to heterogeneity of variance (Levene's p = 0.01), as shown in Table 25, 26 and figure 15, we used Welch's ANOVA for LM (F (3, ≈107.41) = 16.948, p = 0.001), followed by Games–Howell. Both EG1 and EG2 outperformed their respective controls: EG1 vs CG1 (ΔM = 14.76, 95% CI [7.31, 22.21]); EG1 vs CG2 (ΔM = 19.30, 95% CI [10.60, 28.00]); EG2 vs CG1 (ΔM = 12.68, 95% CI [4.23, 21.13]); EG2 vs CG2 (ΔM = 17.22, 95% CI [7.67, 26.77]). There was no significant difference between EG1 and EG2: EG1 vs EG2 (ΔM = 2.08, 95% CI [–5.52, 9.68]).

Sensitivity Analysis

Table 27 Sensitivity Analysis for Post-Test LM

Z-Score Threshold	Number of Outliers	IDs of Outliers	Post_LM Retained for Analysis	*p*
±3.0	0	None	All 200 Participants	0.001

To assess the robustness of the findings for post-test Post_LM, a sensitivity analysis was performed using a ±3.0 Z-score threshold to detect potential outliers (Table 27). No cases exceeded this threshold, and all 200 participants were retained in the final analysis. Importantly, the group differences remained statistically significant (p = 0.001) even without excluding any data points, thereby reinforcing the stability and reliability of the results.

Interactions over time (only groups with pretest)

Two-Way ANOVA

Table 28 Repeated-measures analysis in pretested groups (EG1, CG1)

Variable	Effect	F	p	Partial η^2	Simple Effects
PF	Group	18.838	< .001	0.088	EG1 vs CG1: ΔM = +14.76 (95% CI [9.11, 20.42])
	Time	0.61	.436	-	
	Group × Time	16.31	.000069	0.229	
MH	Group	24.706	< .001	0.112	EG1 vs CG1: ΔM = +12.68 (95% CI [4.23, 21.13])
	Time	0.356	.552	-	
	Group × Time	24.706	< .001	0.112	
LM	Group	47.698	< .001	0.196	EG1 vs CG1: ΔM = +19.30 (95% CI [10.60, 28.00])
	Time	0.848	.360	-	
	Group × Time	19.91	< .001	0.196	

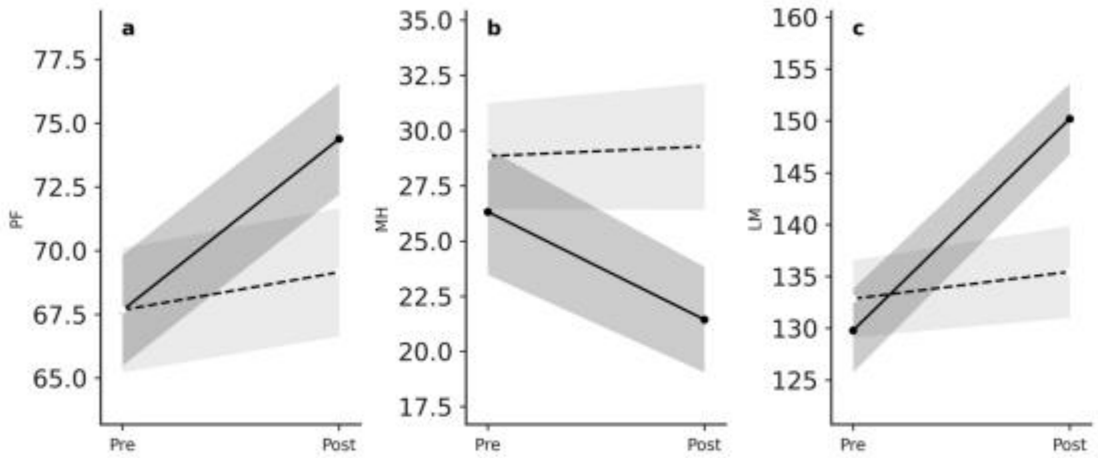


Figure 16 Pre - post trajectories for the pretested groups (EG1 vs CG1)

As shown in Figure 16 and Table 28. A 2 (Group: EG1 vs CG1) × 2 (Time: Pre vs Post) repeated-measures ANOVA was conducted to examine the intervention effects on PF, MH, and LM. For PF, the interaction between group and time reached significance ($F (1, 98) = 16.31$, $p < .001$, $\eta p^2 = .229$), indicating differential changes over time across groups. Within-group change in the EG1 was $\Delta M = +14.76$ (95% CI [9.11, 20.42]).

Regarding MH, a significant Group × Time interaction was also observed ($F (1, 98) = 24.71$, $p < .001$, $\eta p^2 = .112$). The simple effect analysis showed that EG1 exhibited a change of $\Delta M = +12.68$ (95% CI [4.23, 21.13]) from pre- to post-test.

For LM, the Group × Time interaction remained significant ($F (1, 98) = 19.91$, $p < .001$, $\eta p^2 = .196$), with EG1 showing a pre-post difference of $\Delta M = +19.30$ (95% CI [10.60, 28.00]).

Within-Group Effects

Experimental Group (Pre vs Post)

Table 29 Paired Samples t-Test Results for Experimental Group with Pretest

Paired Variables	*Mean Difference*	*Std. Deviation*	*Std. Error Mean*	*95% CI Lower*	*95% CI Upper*	*t*	*df*	*p* (2-tailed)
Pre_PF-Post_PF	-6.716	11.3664	1.6074	-9.9463	-3.4857	-4.18	49	0
Pre_MH - Post_MH	4.88	12.715	1.798	1.267	8.493	2.71	49	0.009
Pre_LM - Post_LM	-20.42	19.396	2.743	-25.932	-14.908	-7.44	49	0

To assess the within-group effect of the GL intervention in EG1, paired samples t-tests were conducted comparing pretest and posttest scores. Significant improvements were observed across all three outcome variables, as shown in Table 29.

For PF, a significant difference was found between pretest and posttest scores, with a mean difference of -6.716 (negative value indicating higher posttest scores), t (49) = -4.18, $p < .001$, demonstrating enhanced physical performance.

In terms of MH, scores decreased by an average of 4.88 points from pre- to post-test (t (49) = 2.71, $p = .009$), where lower scores denote reduced psychological distress and better well-being, indicating a significant positive effect.

The most pronounced change occurred in LM, with posttest scores exceeding pretest scores by an average of 20.42 points, t (49) = -7.44, $p < .001$, strongly supporting the intervention's efficacy in boosting student motivation. The relatively narrow confidence intervals across all variables further affirm the robustness and reliability of these gains. Collectively, these findings provide compelling evidence of the intervention's effectiveness in enhancing physical health, MH, and LM among students.

Control Group (Pre vs Post)

Table 30 Paired Samples t-test Results for Control Group with pretest

Paired Variables	*Mean* Difference	*Std.* Deviation	*Std. Error* Mean	95% *CI* Lower	95% *CI* Upper	*t*	*df*	*p* (2-tailed)
Pre_PF-Post_PF	-1.48	2.27	0.321	-2.1251	-0.8349	-4.61	49	0
Pre_MH - Post_MH	-0.44	13.56	1.918	-4.294	3.414	-0.23	49	0.819
Pre_LM - Post_LM	-2.6	20.448	2.892	-8.411	3.211	-0.90	49	0.373

Table 30 shows that the paired sample t-test results for CG1 show no substantial changes in most of the variables except for a small but statistically significant improvement in PF. Specifically, PF showed a mean difference of -1.48 (negative value indicating higher post-test scores), t (49) = -4.61, p < .001. As shown in Table 30, this improvement, while statistically significant, is modest in magnitude and likely attributable to factors such as test familiarity, natural physical variation, or environmental conditions rather than any structured intervention.

For MH, the mean difference was only -0.44 and not statistically significant (t (49) = -0.23, p = 0.819), suggesting no meaningful change in MH. Similarly, LM did not exhibit significant improvement, with a mean difference of -2.6 (t (49) = -0.90, p = 0.373).

These findings demonstrate that in the absence of the intervention, CG1 experienced no significant shifts in MH or LM, and the modest PF improvement is unlikely to be practically meaningful. Consequently, the marked gains observed in the experimental groups can be more confidently attributed to the intervention rather than to natural development or testing effects.

Subscale-Level Findings

Table 31 PF Seven Project, MH Three Dimensions, LM Eight Dimensions

Variable	Dimensions	Experimental Group		Control group		*F*	*p*
		EG1 (*n* = 50)	EG2 (*n* = 50)	CG1 (*n* = 50)	CG2 (*n* = 50)		
PF	Post-test BMI	14.10 ± 1.52	13.98 ± 1.44	13.86 ± 1.59	13.92 ± 1.69	0.22	0.885
	Post-test VC	11.59 ± 2.06	11.46 ± 1.71	9.73 ± 3.47	9.43 ± 3.95	7.32	<0.001
	Post-test SLJ	7.66 ± 1.23	7.81 ± 1.02	6.41 ± 2.61	6.84 ± 2.12	6.41	<0.001
	Post-test SR	7.72 ± 1.13	7.64 ± 1.32	6.61 ± 2.11	6.72 ± 2.23	5.55	0.001
	Post-test PUSU	7.52 ± 1.60	7.55 ± 1.33	6.43 ± 2.22	6.00 ± 2.61	7.61	<0.001
	Post-test 50Ms	15.04 ± 2.84	14.71 ± 2.74	13.50 ± 4.75	12.05 ± 5.02	5.82	<0.001
	Post-test 8/1Ms	15.17 ± 3.25	15.15 ± 2.16	12.22 ± 5.30	12.61 ± 5.19	7.23	<0.001
MH	Post-test A	4.24 ± 3.42	4.88 ± 5.00	5.64 ± 4.09	5.24 ± 3.70	1.05	0.371
	Post-test D	5.32 ± 4.31	5.72 ± 4.80	6.88 ± 6.74	7.44 ± 4.88	1.76	0.156
	Post-test S	11.88 ± 6.89	11.40 ± 6.78	16.76 ± 5.90	14.48 ± 7.17	6.89	<0.001
LM	Post-test M	16.94 ± 5.59	17.06 ± 5.77	15.80 ± 5.05	15.58 ± 6.43	0.89	0.449
	Post-test En	22.72 ± 3.26	22.40 ± 3.57	18.88 ± 4.93	18.70 ± 5.65	11.97	<0.001
	Post-test Psy	22.64 ± 3.21	23.42 ± 2.80	19.72 ± 4.99	19.72 ± 5.12	10.83	<0.001
	Post-test Phy	23.04 ± 2.51	22.42 ± 3.31	19.24 ± 5.83	19.26 ± 5.39	10.2	<0.001
LM	Post-test App	17.30 ± 5.43	16.44 ± 6.16	17.46 ± 5.70	15.56 ± 5.50	1.18	0.318
	Post-test O	13.72 ± 6.09	14.06 ± 5.63	13.50 ± 5.49	13.30 ± 6.11	0.16	0.926
	Post-test Aff	16.46 ± 5.16	16.66 ± 5.87	15.56 ± 6.05	13.94 ± 6.30	2.24	0.085
	Post-test C	17.38 ± 6.36	15.66 ± 5.98	15.28 ± 6.78	14.84 ± 6.41	1.51	0.212

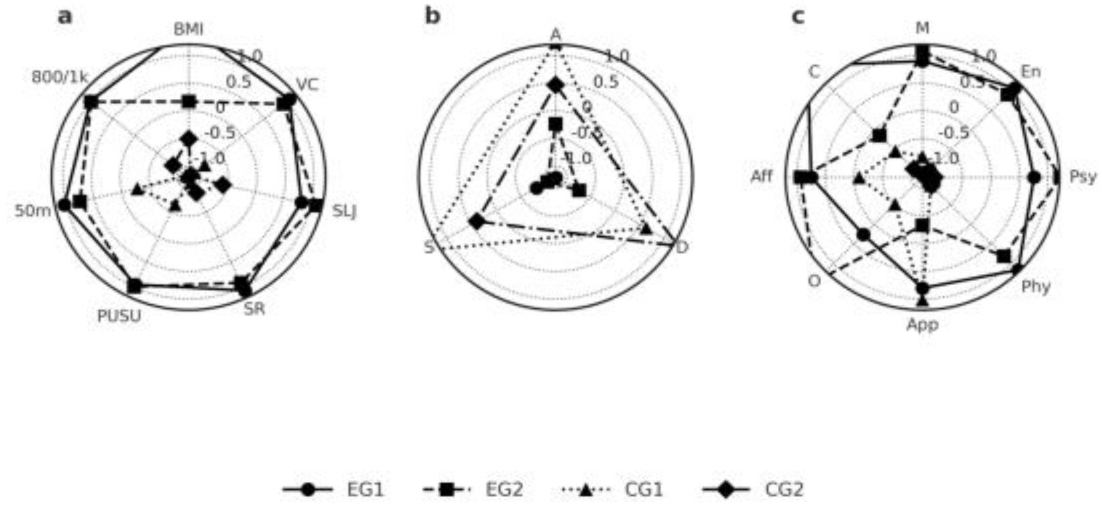


Figure 17 Subscale overview (radar plots, post-test)

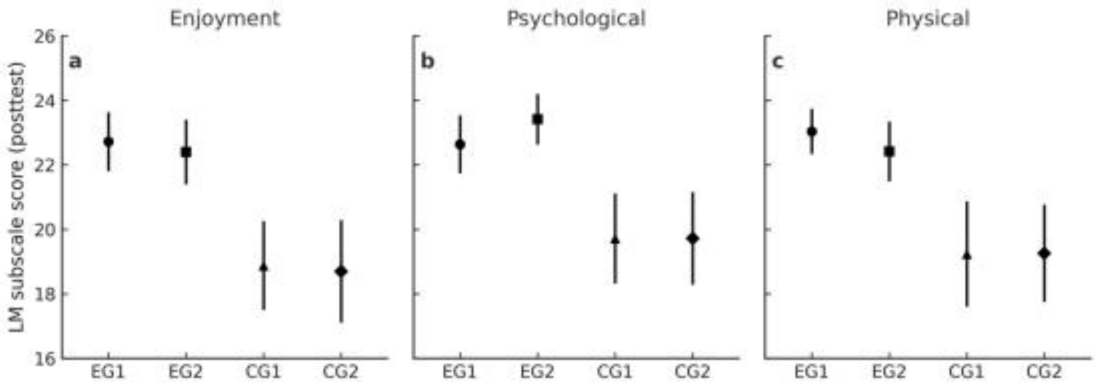


Figure 18 Posttest means (±95% CI) for the three key PALMS subscales of LM

Table 31 and Figure 17 summarize the post-test differences across four Solomon groups in PF 7 indicators, MH 3 dimensions, and LM 8 dimensions, using one-way ANOVA for each outcome.

For PF, all indicators showed significant group differences (p < .01), except BMI (p = .885). In particular, the experimental groups outperformed control groups in vital capacity, standing long jump, sit-and-reach, PUSU, 50-meter sprint, and 800/1000-meter run, suggesting robust intervention effects on cardiopulmonary function, flexibility, muscular strength, and endurance. The non-significant BMI difference may reflect the relatively short intervention duration.

In the MH dimensions, only stress showed a significant group effect (p < .001), with lower stress levels observed in the experimental groups. No significant differences were found in anxiety and depression scores (p > .05), suggesting the intervention's primary psychological benefit was in stress alleviation rather than emotional symptom reduction.

Regarding LM, as shown in Table 32 and Figure 18 significant differences were observed in enjoyment, psychological condition, and physical condition (all p < .001), indicating these facets of motivation were positively influenced by the gamified intervention. Other dimensions appearance, others' expectations, affiliation, and competition/ego did not differ significantly between groups (p > .05), possibly due to their lower relevance to the intervention context.

Between-Subjects Effects of Group, Gender, and Major on Post-Test Outcomes (PF, MH, LM)

Table 32 Between-Subjects Effects of Group, Gender, and Major on Post-test Outcomes (PF, MH, LM)

Post Variable	Source	*F*	*df*	*p*	η^2
Post_PF	Group	4.108	(3,176)	0.008	0.065
Post_MH	Gender	0.297	(1,176)	0.587	0.002
Post_LM	Major	0.63	(2,176)	0.534	0.007
Post_PF	Group × Gender	0.829	(3,176)	0.479	0.014

Table 32 (Continued)

Post Variable	Source	*F*	*df*	*p*	η^2
Post_MH	Group × Major	0.547	(6,176)	0.772	0.018
Post_LM	Gender × Major	0.369	(2,176)	0.692	0.004
Post_PF	Group × Gender × Major	0.978	(6,176)	0.442	0.032
Post_MH	Group	5.869	(3,176)	0.001	0.091
Post_LM	Gender	5.117	(1,176)	0.025	0.028
Post_PF	Major	0.096	(2,176)	0.908	0.001
Post_MH	Group × Gender	0.961	(3,176)	0.413	0.016
Post_LM	Group × Major	1.936	(6,176)	0.077	0.062
Post_PF	Gender × Major	0.984	(2,176)	0.376	0.011
Post_MH	Group × Gender × Major	0.903	(6,176)	0.494	0.03
Post_LM	Group	15.97	(3,176)	0	0.214
Post_PF	Gender	1.015	(1,176)	0.315	0.006
Post_MH	Major	1.471	(2,176)	0.233	0.016
Post_LM	Group × Gender	1.748	(3,176)	0.159	0.029
Post_PF	Group × Major	0.742	(6,176)	0.617	0.025
Post_MH	Gender × Major	1.058	(2,176)	0.349	0.012
Post_LM	Group × Gender × Major	0.644	(6,176)	0.695	0.021

As shown in Table 32, the results from the between-subjects ANOVA showed significant effects of the Group on all three outcomes: PF, MH, and LM. Specifically, significant group differences were observed for PF (F (3,176) = 4.108, p = 0.008, η^2 = 0.065), MH (F (3,176) = 5.869, p = 0.001, η^2 = 0.091), and LM (F (3,176) = 15.97, p < 0.001, η^2 = 0.214). The largest effect size was observed for LM, indicating that the intervention had a substantial impact on LM, with experimental groups outperforming the control groups.

In terms of Gender, a significant effect was found only for LM (F (1,176) = 5.117, p = 0.025, η^2 = 0.028), suggesting a minor gender effect on LM. No significant gender effects were observed for PF or MH.

Regarding Major, no significant effects were found for any of the three outcomes (PF, MH, and LM), indicating that the intervention’s impact did not vary by academic major.

None of the interaction effects (Group × Gender, Group × Major, and Group × Gender × Major) were significant ($p > 0.05$), suggesting that the intervention's effects were stable across different gender and major subgroups.

Game Score Effect and (Exploratory) Mediation Analyses

EG1 vs EG2 Game Score

Table 33 Comparison of Game Scores Between EG1 and EG2 (Independent Samples t-test)

Variable	Group	*N*	*Mean*	*SD*	*SE*	*t*	*df*	*Sig.* (2-tailed)
Score	EG1	50	44.88	7.644	1.081	0.317	98	0.752
	EG2	50	44.26	11.542	1.632			

To assess whether the presence of a pretest influenced game performance, an independent samples t-test was conducted to compare the game scores between EG1 and EG2. As shown in Table 33, the mean score for EG1 was 44.88 (*SD* = 7.64), while EG2 scored a mean of 44.26 (*SD* = 11.54), reflecting only a minimal difference between the two groups.

The *t*-test results showed no statistically significant difference in game scores between EG1 and EG2 ($t = 0.317$, $df = 98$, $p = 0.752$). This suggests that the presence or absence of a pretest did not materially affect students' engagement or performance during the gamified intervention. In practical terms, students’ responses to the gamification were similar regardless of their pretest status, which supports the validity of subsequent analyses across the entire experimental cohort.

Correlation with Post Measures

Table 34 Normality Test Results for Score and Post-test Variables (EG1 and EG2)

Variable	Skewness	Kurtosis	Shapiro-Wilk *W*	Shapiro-Wilk *p*
Score (EG1)	-0.395	0.499	0.979	0.495
Score (EG2)	0.077	0.428	0.988	0.894
PF (EG1)	-0.04	0.666	0.984	0.74
PF (EG2)	-0.488	0.048	0.965	0.142
MH (EG1)	0.045	0.599	0.982	0.642
MH (EG2)	0.164	-0.108	0.99	0.937
LM (EG1)	-0.204	-0.444	0.986	0.815
LM (EG2)	-0.29	-0.573	0.98	0.569

As shown in Table 34, since all *p*-values were higher than the 0.05 critical value, the Shapiro-Wilk test results for normality indicate that the distributions of game scores and post-test variables (Post_PF, Post_MH, and Post_LM) for both EG1 and EG2 closely align with normal distributions. Furthermore, the normality assumption for parametric testing was confirmed by skewness and kurtosis values falling within acceptable ranges.

Before conducting regression and mediation analyses, normality assumptions were assessed for key variables—specifically, game score (Score) and post-test outcomes (Post_PF, Post_MH, Post_LM)—within both EG1 and EG2. Table 34 summarizes the results of skewness, kurtosis, and the Shapiro-Wilk test.

In EG1, skewness values ranged from -0.40 to 0.05, and kurtosis values ranged from -0.44 to 0.67, with all Shapiro-Wilk p-values exceeding 0.05, indicating no significant deviations from normality. Similarly, the variables in EG2 also demonstrated characteristics of normal distribution, with Shapiro-Wilk p-values for Score (p = .894), Post_PF (p = .142), Post_MH (p = .937), and Post_LM (p = .569) all showing non-significance.

Table 35 Pearson Correlation Between Game Score and Post-Intervention Outcomes of EG1 and EG2

Variable	Game score		
Score	1		
PF	-0.439**	1	
MH	0.504**	-0.04	1
LM	0.853**	-0.092	0.238*

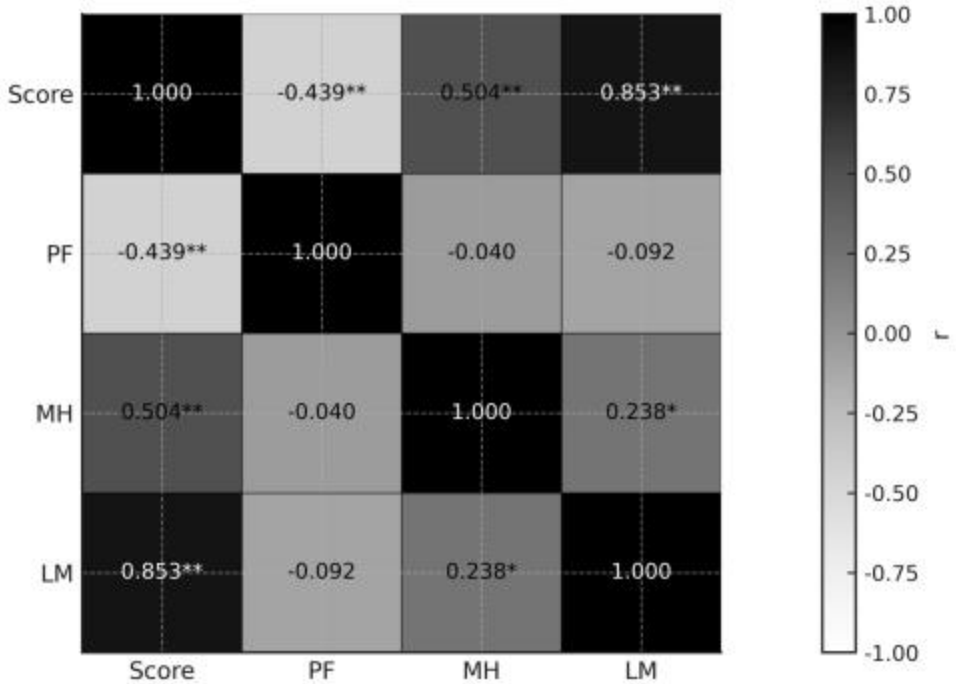


Figure 19 Correlation matrix (post-test, experimental cohorts)

Pearson correlation analysis revealed significant relationships between the game score and post-intervention variables. Specifically, the game score showed a significant positive correlation with Post_LM ($r = 0.853$, $p < 0.01$), a significant negative correlation with Post_PF ($r = -0.439$, $p < 0.01$), and a significant positive correlation with

Post_MH (r = 0.504, p < 0.01), See Figure 19 and Table 35. These results suggest that while higher PF scores are inversely correlated with game scores, possibly reflecting a trade-off in how participants allocate effort or prioritize tasks, higher motivation for learning and better MH are associated with better performance in the game.

To further explore how game performance relates to post-intervention outcomes, Pearson correlation analysis was conducted between game scores and post-test variable Post_PF, Post_MH, and Post_LM in the experimental groups (EG1 and EG2). The results revealed a strong and statistically significant positive correlation between game score and LM (r = 0.853, p < 0.01), suggesting that participants who were more engaged and performed better in the game also reported higher motivation levels after the intervention.

Additionally, a moderate positive correlation was found between game score and MH (r = 0.504, p < 0.01), implying that higher game engagement might be linked to improved MH. In contrast, a significant negative correlation was observed between game score and PF (r = -0.439, p < 0.01), which may reflect an inverse relationship likely influenced by the game design, where cognitive or strategic engagement is emphasized over physical exertion.

Regression Analysis

Table 36 Hierarchical Regression Analysis Predicting LM (Post_LM)

Term	Model 1					Model 2				
	B	*SE*	*t*	*p*	**β**	*B*	*SE*	*t*	*p*	**β**
Constant	150.844	12.318	12.246	0	-	29.587	5.284	5.6	0	
Post_PF	-0.132	0.157	-0.843	0.401	-0.083	0.688	0.053	12.94	0	0.431
Post_MH	0.373	0.157	2.383	0.019	0.234	-0.579	0.055	-10.47	0	-0.363
Score						1.821	0.057	31.75	0	1.225
R^2			0.063					0.919		
AD. R^2			0.044					0.916		
F			3.28					360.878		
ΔR^2								0.855		
ΔF								1007.982		

Note: Dependent variable = Post_LM, * p<0.05 ** p<0.01, AD.R^2=Adjusted R^2, SE=Standard Error

To validate whether the inclusion of game score improves the prediction of LM, two hierarchical regression models were compared. Model 1-2 included only two predictors, As shown in Table 36: post-intervention Post_PF and Post_MH. The results indicated a non-significant model (*F* (2,197) = 0.099, *p* = 0.906), with minimal explanatory power (R^2 = 0.001, Adjusted R^2 = -0.009). Neither of the predictors was statistically significant (Post_PF: β = 0.03, Post_MH: β = -0.006), indicating the inadequacy of this model in predicting LM without accounting for game performance.

In contrast, Model 2 included game score as a third predictor. The inclusion of this variable markedly enhanced model performance, with R^2 increasing to 0.919 and adjusted R^2 reaching 0.916. The model was highly significant (*F* (3,96) = 360.878, $p < 0.001$). Most notably, game score emerged as the strongest predictor (β = 1.225, *t* = 31.749, $p < 0.001$), demonstrating overwhelming influence on post-test LM. Interestingly, the direction of MH's predictive role reversed and became significantly negative (β = -0.363, $p < 0.001$) in the presence of game score, suggesting a potential suppression or mediation effect that warrants further theoretical exploration.

Post-intervention Post_LM was assessed using hierarchical multiple regression. Model 1 revealed a minor but statistically significant amount of variance in Post_LM (R^2 = 0.063, Adjusted R^2 = 0.044, *F* = 3.28, $p < 0.05$), with Post_MH being the only significant positive predictor (*B* = 0.373, *p* = 0.019). Post_PF did not significantly contribute to predicting LM. When game score was added to Model 2, a significant increase in explained variance was observed (R^2 = 0.919; Adjusted R^2 = 0.916; ΔR^2 = 0.855; ΔF = 1007.982; $p < 0.001$), indicating that game score was a powerful predictor of Post_LM (*B* = 1.821, β = 1.225, $p < 0.001$). Notably, the relationship between Post_PF, Post_MH, and Post_LM shifted when game score was added to the model. Post_PF became a significant positive predictor (*B* =.688), while Post_MH became a significant negative predictor (*B* = -0.579). These modifications suggest that game score may moderate or mitigate the impacts of Post_PF and Post_MH as predictors of LM. According to the final model, game performance was the primary factor influencing the results.

High vs Low Score Group

Table 37 Normality and Homogeneity Tests for Post-Test Variables by Score Group

Variable	Shapiro-Wilk p (High Score)	Shapiro-Wilk p (Low Score)	Levene's Test F	Levene's Test p
Post_PF	0.215	0.143	1.34	0.25
Post_MH	0.488	0.487	0.353	0.554
Post_LM	0.575	0.904	2.137	0.147

Prior to conducting the comparative analysis based on game score groups (high vs. low), this study examined whether the post-intervention variables met the statistical assumptions of normality and homogeneity of variance. This step ensures the validity of subsequent parametric testing. As shown in Table 37, the Shapiro-Wilk test results indicated that post-test Post_PF, Post_MH and Post_LM were normally distributed in both the high score group (*p* = 0.215, 0.143, 0.575) and the low score group (*p* = 0.143, 0.48, 0.904).

Levene's test results showed no significant differences in variance between the two groups for any of the three outcome variables. Specifically, F-values for Post_PF (1.34), Post_MH (0.353), and Post_LM (2.137) were all nonsignificant ($p > 0.05$), indicating that the assumption of homogeneity of variance was satisfied.

These findings affirm that the data distributions across the high and low score groups were statistically comparable, allowing for the appropriate application of parametric inferential tests in subsequent comparisons. Therefore, the satisfactory results of normality and homogeneity tests ensure the statistical validity and reliability of our analyses.

Table 38 Comparison of Post-Test Outcomes Between High and Low Score Groups (EG1 & EG2)

Variable	High *(Mean±SD)*	Low *(Mean±SD)*	*t*	*p*
Post_PF	70.82 ± 9.50	77.18 ± 7.57	3.714	< .001
Post_MH	25.46 ± 9.01	18.27 ± 7.76	-4.285	< .001
Post_LM	158.50 ± 10.04	140.54 ± 12.51	-7.879	< .001

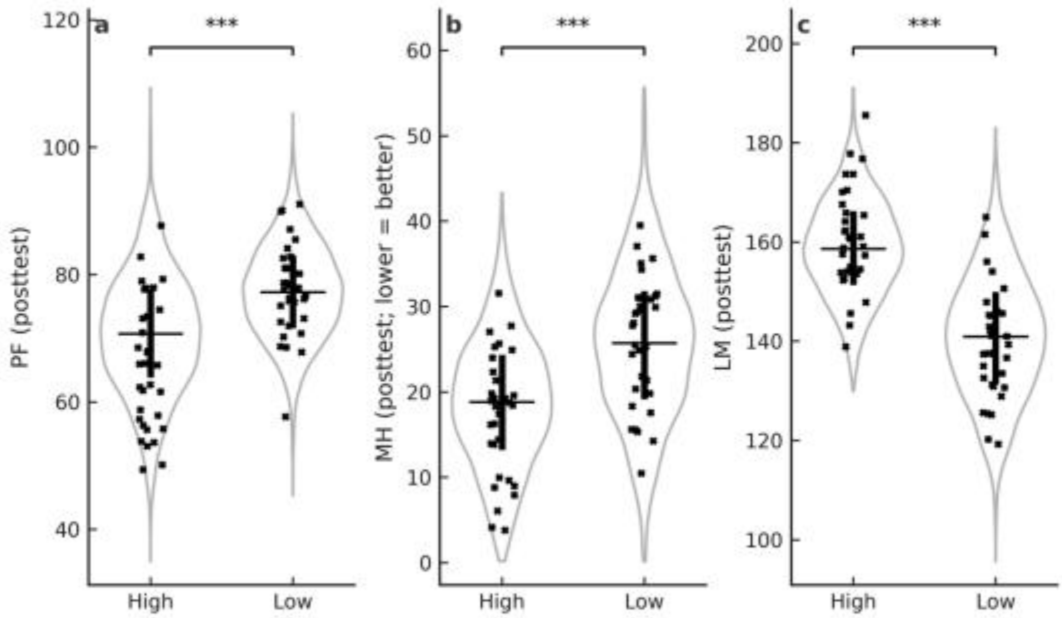


Figure 20 High vs Low (Score) comparisons.

According to the normality and homogeneity tests, all post-test variables were found to be normally distributed in both the high and low score groups, with Shapiro-Wilk p-values greater than 0.05. Additionally, Levene's tests revealed no significant differences between the groups, confirming homogeneity of variance for Post_PF (*p* = 0.25), Post_MH (*p* = 0.554), and Post_LM (*p* = 0.147). Thus, independent samples t-tests were conducted under these assumptions. For all outcomes, the differences between the high and low score groups were found to be highly significant.

Interestingly, As shown in Figure 20 and Table 38, the results showed that the low score group outperformed the high score group in post-test PF (*M* = 77.18 vs. 70.82, *t* = 3.714, *p* < .001). This unexpected trend may suggest that students in the high score group exerted more energy on competitive gaming components, which could have led to fatigue or uneven physical performance. In contrast, the high score group exhibited significantly better MH outcomes, with lower stress and negative emotional states (M = 18.27 vs. 25.46, *t* = -4.285, *p* < .001), indicating a more favorable MH adjustment during the intervention.

Most notably, LM scores were substantially higher in the high score group (*M* = 158.50 vs. 140.54, *t* = -7.879, *p* < .001), emphasizing the motivational effectiveness of the gamified scoring system.

Taken together, these results highlight that while game scores may not directly translate into better PF, they are closely linked with improvements in MH and LM, reinforcing the nuanced impact of gamification in educational interventions.

Exploratory SEM: The Mediating Role of Game Score in LM

Table 39 Mediation Analysis: Game Score as a Mediator in the Relationship Between MH and LM

Metric/Predictor	Coefficient (b)	*SE*	*t*	*p*	95% *CI*
Model Summary					
R	0.66				
R-squared	0.43				
F (2, 97)	36.64			0	
Model Coefficients					
constant	66.58	6.46	10.3	0	[53.76, 79.41]
Post_MH	0.52	0.08	6.36	0	[0.36, 0.69]
Post_PF	-0.45	0.08	-5.47	0	[-0.61, -0.29]

As shown in Table 39, the mediating model examines whether game performance scores mediate the relationship between MH and LM, while controlling for physical health variables. The overall model performs well, F (2,97) = 36.64, $p < .001$, $R^2 = 0.43$, indicating that the model explains 43% of the variance in LM. A significant positive relationship is found between MH and LM, with a coefficient of 0.52 ($p < .001$, 95% *CI* [0.36, 0.69]). However, an unexpected negative correlation is observed between physical health and LM, with a coefficient of -0.45 ($p < .001$, 95% *CI* [-0.61, -0.29]).

CHAPTER 5
DISCUSSION AND CONCLUSION

DISCUSSION

Opening Synthesis: What Changed and How Much

Using a Solomon four-group experimental design, this GL intervention produced moderate improvements in PF (Cohen's d = 0.61), large gains in LM (Cohen's *d* = 0.97), and moderate-to-large reductions in overall MH distress (Cohen's d = 0.70) among higher vocational students in health-oriented programs. However, when examining MH subdomains, only stress showed significant improvement (Cohen's d = 0.59), while anxiety (p = .371) and depression (p = .156) showed no significant changes. Additionally, BMI remained unchanged (p = .885).

Importantly, the Solomon design confirmed no pretest sensitization effects. All Treatment × Pretest interactions remained non-significant (all $p > .15$), with significant treatment main effects across all outcomes (PF: ηp^2 = .088, $p < .001$; MH: ηp^2 = .112; LM: ηp^2 = .196, $p < .001$), establishing internal validity and confirming that observed changes resulted from the intervention rather than testing artifacts.

The effect size hierarchy (LM > MH > PF) aligns with recent gamification meta-analyses in educational contexts, which consistently report stronger effects on motivational outcomes compared to behavioral or physiological measures (Dichev & Dicheva, 2017; Koivisto & Hamari, 2019). This pattern suggests that gamification primarily operates through psychological mechanisms that subsequently influence behavioral and physical outcomes through mediated pathways.

Theoretical Interpretation: SDT and Gamification Mechanisms

The differential effect patterns across outcomes strongly support SDT as the primary explanatory framework for gamification effectiveness in educational contexts. SDT posits that motivation and well-being are enhanced when interventions satisfy three basic psychological needs: autonomy, competence, and relatedness (Deci & Ryan, 2012). The observed effect size hierarchy directly corresponds to the directness of psychological need satisfaction provided by the gamification mechanisms, demonstrating SDT's critical role in understanding how gamified learning influences PF, MH, and LM.

SDT and LM Enhancement

The large LM effects (d = 0.97) reflect the gamification system's strong support for autonomy needs through choice provision, multiple success pathways, and self-paced progression. The strong correlation between game performance and LM (r = 0.853, $p < .001$, as reported in Table 35) indicates that students who experienced greater agency developed stronger intrinsic motivation, consistent with recent meta-analytic evidence showing autonomy support as the strongest predictor of intrinsic motivation in educational settings (Vasconcellos et al., 2020).

The LM improvements were selective, with significant effects observed in enjoyment ($p < .001$), MH ($p < .001$), and PF ($p < .001$), while extrinsic factors showed no significant changes (appearance: $p = .318$; others' expectations: $p = .926$). This pattern aligns with SDT's prediction that autonomy-supportive environments primarily enhance self-determined motivation rather than controlled forms of motivation (Ntoumanis et al., 2021). The gamification system fulfilled students' need for autonomy by providing choices and self-direction, which directly translated into enhanced intrinsic motivation for physical education participation.

SDT and PF Development

The hierarchical regression analysis revealed game score as the strongest predictor of LM (β = 1.225, $p < .001$, explaining an additional R^2 change of 0.855 as shown in Table 36), indicating that competence experiences within the gamified system were the primary drivers of motivational enhancement. From an SDT perspective, this demonstrates how satisfaction of the competence need through mastery experiences and achievement progression creates a foundation for both motivational and physical outcomes.

The moderate PF improvements (d = 0.61) likely resulted from competence-mediated behavioral engagement: enhanced self-efficacy beliefs ⟶ increased physical activity participation ⟶ improved fitness outcomes. This mediation pathway is consistent with social cognitive theory's emphasis on self-efficacy as the primary mediator between environmental supports and behavioral change (Bandura, 2006; Deci et al., 2017).

Component analyses revealed differential responses across PF measures, all interpreted through SDT's lens of competence building. Significant improvements were observed in cardiovascular endurance (vital capacity: $p < .001$), muscular endurance (pull-ups/sit-ups: $p < .001$), explosive power (standing long jump: $p < .001$), and aerobic capacity (800/1000m runs: $p < .001$). These improvements reflect how competence need satisfaction through progressive challenges and immediate feedback motivated students to engage more intensively in physical activities, leading to rapid neuromuscular and cardiovascular adaptations. The gamification system's competence-supportive features, clear goals, immediate feedback, and visible progress, aligned with SDT's emphasis on mastery experiences as drivers of sustained behavioral engagement.

The absence of BMI changes ($p = .885$) aligns with SDT's recognition that while psychological need satisfaction can drive behavioral engagement, structural physiological changes require sustained energy balance modifications over longer periods. The observed pattern, functional improvements without body composition changes, demonstrates that SDT mechanisms primarily influence behavioral engagement and skill development rather than metabolic adaptations requiring extended intervention durations.

SDT and MH Outcomes

The selective stress reduction ($d = 0.59$, $p < .001$) without changes in anxiety ($p = .371$) or depression ($p = .156$) can be interpreted through SDT's relatedness need and its differential impact on various mental health dimensions. The gamification system primarily addressed social-evaluative stressors through relatedness satisfaction: team-based competition, collaborative tasks, and shared goal pursuit fostered social connection while reducing social evaluation apprehension common in traditional PE contexts.

From an SDT perspective, the stress improvements reflect how relatedness need satisfaction creates a supportive social environment that reduces situational stress appraisals. The gamified intervention directly addressed primary stress-generating factors in physical education: unpredictable evaluation criteria (autonomy threat), public performance pressure (competence threat), and lack of social support (relatedness threat). By satisfying all three basic psychological needs, the intervention reduced mean stress scores from approximately 15.62 in controls to 11.64 in experimental groups, representing meaningful practical differences in students' daily stress experience.

However, anxiety and depression, which typically involve more stable cognitive-emotional patterns, proved less responsive to the 16-week intervention focused on need satisfaction in the PE context. SDT acknowledges that while basic psychological need satisfaction promotes general well-being, clinical symptoms may require more targeted therapeutic approaches beyond environmental restructuring (Deci et al., 2017). The baseline scores were relatively low across all groups, with high individual variability, suggesting that SDT-based interventions may be most effective for preventing deterioration and addressing situational distress rather than treating established psychological disorders.

The mediation analysis provided preliminary evidence for SDT's proposed pathway linking psychological need satisfaction to multiple outcomes. The model examining the relationship between MH and LM through game performance explained 43% of the variance in LM (R^2 = 0.43, Table 39), with MH serving as a significant predictor of game score (b = 0.52, $p < .001$). This supports SDT's proposition that well-being (reflected in reduced MH distress) and competence experiences (reflected in game performance) jointly contribute to enhanced motivation. However, the moderate explained variance suggests that additional unmeasured factors, potentially including external life stressors, prior physical education experiences, and peer relationship quality, also substantially contribute to LM outcomes, indicating areas for future SDT-based research.

Integration: SDT's Comprehensive Impact Across Outcomes

The differential effects across PF, mental health, and learning motivation provide strong empirical support for SDT's theoretical framework while revealing important nuances. The effect size hierarchy (LM > MH > PF) directly reflects the proximity of outcomes to psychological need satisfaction: learning motivation, being fundamentally psychological, responds most directly to need satisfaction; mental health (specifically stress) benefits from need satisfaction but is constrained by intervention duration and symptom stability; PF improvements require sustained behavioral engagement driven by need satisfaction but also depend on physiological adaptation timeframes.

SDT's three basic psychological needs (autonomy, competence, and relatedness) operated synergistically within the gamified system. Autonomy support through choice and self-paced progression enhanced intrinsic motivation. Competence building through progressive challenges and immediate feedback increased self-efficacy and behavioral engagement. Relatedness satisfaction through collaborative activities and supportive competition reduced social stress. This multi-faceted need satisfaction explains why gamification produced broad benefits across motivational, psychological, and physical domains.

The study demonstrates that SDT-based gamification design can systematically address the triple crisis facing higher vocational education students through three pathways: motivational deficits are addressed through autonomy and competence support, mental health challenges through relatedness and stress reduction, and PF decline through competence-mediated behavioral engagement. The selective nature of improvements (significant effects for intrinsic motivation factors, performance-based fitness measures, and situational stress, but not for extrinsic motivation, body composition, or clinical psychological symptoms) reveals both the strengths and boundaries of SDT-based educational interventions.

Methodological Strengths and Limitations

Experimental Design Validation

The Solomon four-group design successfully addressed the primary threat to internal validity in educational intervention research: pretest sensitization. The 2 X 2 ANOVA results confirmed no significant pretest main effects (PF: p = .436; MH: p = .552; LM: p = .091) and no Treatment X Pretest interactions (PF: p = .688; MH: p = .307; LM: p = .393), confirming that observed effects were not artifacts of baseline testing procedures.

Baseline demographic equivalence was established across gender distribution (p = .982), academic major (p = .294), and age (p = .425). Outcome measure equivalence in pretested groups was confirmed with no significant differences in PF (p = .998), MH (p = .188), or LM (p = .277).

Statistical Assumption Management

Variance heterogeneity in LM required appropriate statistical adjustment. Levene's test revealed significant variance heterogeneity (p = .01), necessitating Welch's ANOVA (p = .001) and Games-Howell post hoc procedures to maintain appropriate Type I error control. This methodological rigor strengthens confidence in the LM findings.

Normality assumptions were satisfied for all variables based on Shapiro-Wilk tests (all $p > .05$), supporting the use of parametric statistical procedures.

Sample and Duration Limitations

Context-Specific Sample. The sample restriction to health-oriented vocational programs presents both advantages and constraints. The demographic characteristics (mean age ≈ 19.6 years, 64-66% female) define a specific population for which findings are most directly applicable, but limit generalizability to other educational contexts.

Temporal Constraints. The 16-week intervention period was adequate for detecting performance-based fitness improvements and motivational changes but insufficient for body composition alterations or long-term behavioral maintenance assessment. The retention rate of 98.5% (197/200 participants) minimized attrition bias, but the brief follow-up period constrains conclusions about sustained effectiveness.

Implications and Future Directions

Practical Implementation Guidelines

Personalized Challenge Systems. The strong correlation between game performance and LM (r = 0.853) indicates that maintaining optimal individual challenge levels is crucial for competence need satisfaction. Implementation should incorporate adaptive difficulty algorithms that adjust challenge parameters based on real-time performance data to ensure each student experiences appropriate levels of challenge that foster competence without causing frustration.

Immediate Feedback Integration. The regression analysis showing game score as the primary predictor of LM (β = 1.225) emphasizes the importance of immediate, specific performance feedback for competence need satisfaction. Educational implementations should prioritize systems providing instantaneous feedback rather than delayed evaluation, as immediate feedback directly supports students' sense of competence and progress.

Recovery Management. The negative correlation between game scores and PF measures (r = -0.439, $p < .01$, as shown in Table 35) suggests potential resource allocation effects. Implementation protocols should incorporate adequate recovery periods to prevent fatigue that might compromise high-intensity physical performance while maintaining psychological engagement.

Stress-Focused Interventions. The selective stress reduction without anxiety/depression changes suggests interventions could be enhanced through explicit stress management training integrated within the gamified framework, complementing SDT's relatedness support with targeted coping skill development.

Research Priorities

Extended SDT Mechanism Testing. Future studies should employ longitudinal designs with multiple measurement points to examine temporal dynamics of psychological need satisfaction and how sustained need satisfaction translates into stable behavioral and psychological outcomes over 6-12 month periods.

Diverse Population Replication. The context-specific sample necessitates systematic replication across diverse educational settings and student populations to establish generalizability boundaries of SDT-based gamification effects and identify potential moderators.

Objective Measurement Integration. Future research should incorporate accelerometry, heart rate monitoring, and physiological stress markers alongside self-report measures to provide stronger evidence for real-world behavioral change and validate self-reported psychological experiences.

Comprehensive Mediation Models. The exploratory mediation findings require replication through pre-registered studies with a priori hypotheses about specific causal pathways linking gamification elements to psychological need satisfaction and subsequently to physical, psychological, and motivational outcomes.

CONCLUSION

Using a Solomon four-group experimental design, this gamified learning intervention produced moderate improvements in PF ($d = 0.61$), large gains in learning motivation ($d = 0.97$), and moderate-to-large reductions in overall mental health distress ($d = 0.70$) among higher vocational students in health-oriented programs. Importantly, pretest exposure did not alter treatment effects, confirming internal validity.

Component analyses revealed selective benefits and important exceptions. PF improvements were consistent across cardiovascular endurance, muscular endurance, and explosive power measures, whereas body mass index showed no significant change ($p = .885$). Learning motivation improvements were driven by enhanced enjoyment and intrinsic factors, while extrinsic motivators (appearance, others' expectations) remained unchanged. Mental health benefits were limited to stress reduction ($p < .001$), with anxiety ($p = .371$) and depression ($p = .156$) showing no significant changes.

SDT emerged as the primary explanatory framework for understanding these differential effects. The gamification system satisfied students' three basic psychological needs, autonomy, competence, and relatedness, producing cascading effects across outcomes. For learning motivation, autonomy support through choice provision and self-paced progression directly enhanced intrinsic motivation, as evidenced by significant improvements in enjoyment ($p < .001$) and self-determined motivation factors while extrinsic motivators remained unchanged. This aligns with SDT's core prediction that autonomy-supportive environments primarily enhance intrinsic rather than controlled motivation.

For PF, competence need satisfaction through progressive challenges, immediate feedback, and visible achievement progression served as the primary mechanism. The strong predictive relationship between game performance and learning motivation ($\beta = 1.225$, $p < .001$) demonstrates how competence experiences drove both psychological engagement and subsequent behavioral participation. Enhanced self-efficacy beliefs led to increased physical activity engagement, producing measurable improvements in performance-based fitness indicators. However, the absence of body composition changes ($p = .885$) reveals the temporal constraints on physiological adaptations, requiring sustained engagement periods beyond the 16-week intervention.

For mental health, relatedness need satisfaction through team-based activities and collaborative goal pursuit reduced situational stress ($p < .001$) by addressing social-evaluative threats in physical education contexts. The selective stress reduction without anxiety or depression changes demonstrates that SDT-based interventions effectively address situational psychological distress linked to need thwarting but have limited impact on more stable clinical symptoms requiring extended or targeted therapeutic approaches.

The effect size hierarchy (LM > MH > PF) reflects the directness of psychological need satisfaction according to SDT principles: motivation, being fundamentally psychological, responds most directly to need satisfaction; mental health benefits emerge from reduced need thwarting but are constrained by symptom stability; PF improvements require need-driven behavioral engagement plus sufficient physiological adaptation time. The preliminary mediation evidence ($R^2 = 0.43$) suggests that game performance serves as a competence indicator linking need satisfaction to outcomes, though the moderate explained variance indicates additional unmeasured factors warrant investigation.

Given the observed selective improvements, future implementations should incorporate personalized challenge progression to maintain optimal competence experiences, immediate feedback systems to support autonomy and competence needs, adequate recovery scheduling to sustain engagement without overload, and stress-focused components to complement relatedness support. The context-specific sample (health-oriented programs) and brief intervention duration (16 weeks) limit generalizability and preclude conclusions about long-term behavioral maintenance.

These findings support gamified learning as a promising Self-Determination Theory-based approach for enhancing student engagement, reducing stress, and improving fitness in higher vocational education. The study provides empirical evidence linking SDT's psychological need satisfaction to concrete educational outcomes, with specific effect size benchmarks for future intervention development. However, extended intervention periods, diverse population studies, and objective physical activity measurement are needed to establish broader effectiveness, optimal implementation guidelines, and comprehensive understanding of SDT mechanisms in gamified educational contexts. Future research should prioritize longitudinal SDT mechanism testing, diverse population replication, objective measurement integration, and comprehensive mediation models to advance theoretical understanding and practical application of SDT-based gamification in educational settings.

REFERENCES

Adling, R. B., & Chamle, S. S. (2021). Importance of daily physical activity. *International Journal of Health, Physical Education & Computer Science in Sports*, 324.

Ai-jiao, X. (2007). Survey of Depression Situations in Higher Vocational Girl Students—— Taking Hubei Vocational-Technical College As An Example. *Journal of Hubei Vocational-Technical College.*

Ali, A. H., Ahmed, H. S., Jawad, A. S., & Mustafa, M. A. (2021). Endorphin: function and mechanism of action. *Sci Arch*, *2*(1), 9-13.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of personality and social psychology*, *78*(4), 772.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological assessment*, *10*(2), 176.

Armstrong, M. B., & Landers, R. N. (2018). Gamification of employee training and development. *International Journal of Training and Development*, *22*(2), 162-169.

Aston, E. (2018). The Relationships between BMI, Facebook Social Comparison, Body Shape Concerns and Social Physique Anxiety in Female Undergraduates.

Atmaja, P. W., & Mandyartha, E. P. (2020). Gamification of assessment test through multiple question paths to facilitate participants' autonomy and competence. *Letters in Information Technology Education (LITE)*, *3*(1), 9-17.

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., Hwang, I., Kessler, R. C., Liu, H., & Mortier, P. (2016). Mental disorders among college students in the World Health Organization world mental health surveys. *Psychological medicine*, *46*(14), 2955-2970.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents*, *5*(1), 307-337.

Barrett, A., & Muzaffar, T. (2014). Spatial cognitive rehabilitation and motor recovery after stroke. *Current opinion in neurology*, *27*(6), 653-658.

Beets, M. W., & Pitetti, K. H. (2005). Contribution of physical education and sport to health‐related fitness in high school students. *Journal of School Health*, *75*(1), 25-30.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, *37*(1), 179-255.

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented Reality in education–cases, places and potentials. *Educational Media International*, *51*(1), 1-15.

Braver, M. W., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological bulletin*, *104*(1), 150.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour research and therapy*, *35*(1), 79-89.

Budiman, R. D. A., Surjono, H. D., & Hidayat, E. (2025). Gamification: Enhancing Constructivist Skills through Line-Following Robots. *International Journal of Online & Biomedical Engineering*, *21*(1).

Cebrián-CiNotas, S., Ros, C., Fernández-Piqueras, R., & Guerrero, E. (2021). Analysis of the teaching digital skills and the use of ICT resources after a university intervention process, which was based on the implementation of an innovative gamification methodology. *Bordon. Revista de Pedagogia*, *73*(2), 41-61.

Central Committee of the Communist Party of China & State Council. (2016). *Healthy China 2030 Planning Outline*. https://www.gov.cn/zhengce/2016-10/25/content_5124174.htm

Central Committee of the Communist Party of China & State Council. (2019). *China Education Modernization 2035*. S. C. o. t. P. s. R. o. China. https://www.gov.cn/zhengce/2019-02/23/content_5367987.htm

Central Committee of the Communist Party of China, S. C. (2019). *China Education*

Modernization 2035. https://www.gov.cn/zhengce/2019-02/23/content_5367987.htm

Chang, K.-E., Zhang, J., Huang, Y.-S., Liu, T.-C., & Sung, Y.-T. (2020). Applying augmented reality in physical education on motor skills learning. *Interactive Learning Environments*, *28*(6), 685-697. https://doi.org/10.1080/10494820.2019.1636073

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and Technology*, *57*(3), 359-377.

Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of data and information science*, *2*(2), 1-40.

Chen, C., Ibekwe‐SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple‐perspective cocitation analysis. *Journal of the American Society for information Science and Technology*, *61*(7), 1386-1409.

China, M. o. E. o. t. P. s. R. o. (2014). *Notice on issuing the National Student Physical Health Standard (Revised in 2014)*. Ministry of Education of the People's Republic of China. http://www.moe.gov.cn/s78/A17/twys_left/moe_938/moe_792/s3273/201407/t20140708_171692.html

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/https://doi.org/10.4324/9780203771587

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Commission, E. (2019). *Horizon 2020: The EU Framework Programme for Research and Innovation*.

Commission, E. (2020). *Digital Education Action Plan 2021-2027: Resetting education and training for the digital age*. https://ec.europa.eu/education/sites/education/files/document-library-docs/deap-communication-sept2020_en.pdf

Coulter, M., & Ní Chróinín, D. (2013). What is PE? *Sport, Education and Society*, *18*(6), 825-841. https://doi.org/10.1080/13573322.2011.613924

Dan, C. (2007). Function of College Physical Education in Implementing Quality Education. *Journal of Shenyang University*.

de Sousa Mendes, D., de Lima, M. R., de Freitas, R., & Aperecida, T. (2022). Gamification,"I have no idea what it is": a study in the Physical Education Initial Teacher Training. *ALTERIDAD. Revista de Educación*, *17*(1), 12-23.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual review of organizational psychology and organizational behavior*, *4*, 19-43.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. *Handbook of theories of social psychology*, *1*(20), 416-436.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments,

Dettori, J. R. (2011). Loss to follow-up. *Evidence-based spine-care journal*, *2*(01), 7-10.

Dexqonov, B. B. (2024). The role of physical culture and sports in the higher education system. INTERNATIONAL SCIENTIFIC CONFERENCE" INNOVATIVE TRENDS IN SCIENCE, PRACTICE AND EDUCATION",

Dichev, C., & Dicheva, D. (2017). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education*, *14*, 1-36.

Dimitrov, D. M., & Rumrill, J., Phillip D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159-165.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management*, *37*(6), 817-842.

Du, Y. (2023). Exploration of Gamification Teaching Mode in Primary School Mathematics under the Guidance of Core Literacy. *Frontiers in Educational Research*, *6*(24).

Dukh, T., Bodnar, I., Pavlova, I., Svyshch, Y., & Pavlos, O. (2019). Perspectives of using athletics means for improving the level of physical health of students. SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference,

Durão, N. F. S., Moreira, F., Ferreira, M. J. d. S. C., Pereira, C. S., & Annamalai, N. (2020). The state of mobile learning supported by gamification and augmented reality in higher education institutions across three continents. *Revista EDaPECI*, *20*(1), 130-147.

Esra, M., & Sevilen, Ç. (2021). Factors influencing EFL students' motivation in online learning: a qualitative case study. J Educ Technol Online Learn 4 (1): 11–22. In.

Fan, Y., Zheng, H., Ebonite, R. S., De Asis, W. R., & Juanatas, R. A. (2024). Overview and developmental analysis of China's technical and vocational education and training. *International Journal of Innovative Research and Scientific Studies*, *7*(1), 251-260.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fernandez-Rio, J., de las Heras, E., González, T., Trillo, V., & Palomares, J. (2020). Gamification and physical education. Viability and preliminary views from students and teachers. *Physical Education and Sport Pedagogy*, *25*(5), 509-524. https://doi.org/10.1080/17408989.2020.1743253

Fernandez-Rio, J., Zumajo-Flores, M., & Flores-Aguilar, G. (2022). Motivation, basic psychological needs and intention to be physically active after a gamified intervention programme. *European Physical Education Review*, *28*(2), 432-445.

Ferriz-Valero, A., Østerlie, O., García Martínez, S., & García-Jaén, M. (2020). Gamification in Physical Education: Evaluation of Impact on Motivation and Academic Performance within Higher Education. *International Journal of Environmental Research and Public Health*, *17*(12), 4465. https://www.mdpi.com/1660-

4601/17/12/4465

Fewtrell, M. S., Kennedy, K., Singhal, A., Martin, R. M., Ness, A., Hadders-Algra, M., Koletzko, B., & Lucas, A. (2008). How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Archives of disease in childhood*, *93*(6), 458-461.

Filgona, J., Sakiyo, J., Gwany, D. M., & Okoronka, A. U. (2020). Motivation in Learning. *Asian Journal of Education and Social Studies*, *10*(4), 16-37. https://doi.org/10.9734/ajess/2020/v10i430273

Forndran, F., & Zacharias, C. R. (2019). Gamified experimental physics classes: A promising active learning methodology for higher education. *European Journal of Physics*, *40*(4), 045702.

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2015). Toward a new definition of mental health. *World Psychiatry*, *14*(2), 231-233. https://doi.org/10.1002/wps.20231

Gan, X., Wen, X., Lu, Y., & Yu, K. (2019). Economic Growth and Cardiorespiratory Fitness of Children and Adolescents in Urban Areas: A Panel Data Analysis of 27 Provinces in China, 1985–2014. *International Journal of Environmental Research and Public Health*, *16*(19), 3772. https://www.mdpi.com/1660-4601/16/19/3772

Gao, F. (2024). Advancing gamification research and practice with three underexplored ideas in self-determination theory. *TechTrends*, *68*(4), 661-671.

Goldstein, C. E., Weijer, C., Brehaut, J. C., Fergusson, D. A., Grimshaw, J. M., Horn, A. R., & Taljaard, M. (2018). Ethical issues in pragmatic randomized controlled trials: a review of the recent literature identifies gaps in ethical argumentation. *BMC Medical Ethics*, *19*(1), 14. https://doi.org/10.1186/s12910-018-0253-x

Grubic, N., Badovinac, S., & Johri, A. M. (2020). Student mental health in the midst of the COVID-19 pandemic: A call for further research and immediate solutions. *International Journal of Social Psychiatry*, *66*(5), 517-518. https://doi.org/10.1177/0020764020925108

Guo, Y.-f., Liao, M.-q., Cai, W.-l., Yu, X.-x., Li, S.-n., Ke, X.-y., Tan, S.-x., Luo, Z.-y., Cui, Y.-

f., & Wang, Q. (2021). Physical activity, screen exposure and sleep among students during the pandemic of COVID-19. *Scientific reports*, *11*(1), 8529.

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). Multivariate data analysis 7thed Pearson Educational international. *New Jersey*.

Halamish, V., Madmon, I., & Moed, A. (2019). Motivation to learn. *Experimental psychology*.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?--a literature review of empirical studies on gamification. 2014 47th Hawaii international conference on system sciences,

Hammerstein, S., König, C., Dreisörner, T., & Frey, A. (2021). Effects of COVID-19-Related School Closures on Student Achievement-A Systematic Review [Systematic Review]. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.746289

Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (1st ed.). https://doi.org/10.4324/9780203887332

Henry, J. D., & Crawford, J. R. (2005). The short‐form version of the Depression Anxiety Stress Scales (DASS‐21): Construct validity and normative data in a large non‐clinical sample. *British journal of clinical psychology*, *44*(2), 227-239.

Herbert, B., Charles, D., McNeill, M., Moore, A., & Charles, M. (2011). Dynamic Virtual Learning Landscapes to Enhance Student Reflective Processes. European Conference on Games Based Learning,

Holden, M. K. (2005). Virtual environments for motor rehabilitation. *Cyberpsychology & behavior*, *8*(3), 187-211.

Hong-mei, W. (2006). The Investigation of Mental Health of Students in Higher Vocational Colleges. *Journal of Nantong University.*

Hong-yin, W. (2011). Physical Exercise to Promote Physical and Psychological Health of Vocational College Students——For Exampl of Students of Huzhou Vocational and Technological College.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation* *modeling: a multidisciplinary journal*, *6*(1), 1-55.

Hu, T., Zhang, M.-l., Liu, H., Liu, J.-c., Pan, S.-j., Guo, J.-h., Tian, Z.-e., & Cui, L. (2022). The influence of “small private online course + flipped classroom” teaching on physical education students’ learning motivation from the perspective of self-determination theory [Original Research]. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.938426

Jadhav, R. R. (2017). The effect of exercise on anxiety of college students. *International Journal of Physical Education, Sports and Health*, *4*(4), 231-233.

Jagušt, T., Boticki, I., Mornar, V., & So, H.-J. (2017). Gamified digital math lessons for lower primary school students. 2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI),

Jayalath, J., & Esichaikul, V. (2022). Gamification to Enhance Motivation and Engagement in Blended eLearning for Technical and Vocational Education and Training. *Technology, Knowledge and Learning*, *27*(1), 91-118. https://doi.org/10.1007/s10758-020-09466-2

Jensen, C. T., Heinze, C., Andersen, P. K., Bauman, A., & Klinker, C. D. (2022). Mental health and physical activity in vocational education and training schools students: a population-based survey. *European journal of public health*, *32*(2), 233-238.

Jun, R. (2022). Analysis on Characteristics and Innovation of Physical Education Teaching Reform in Higher Vocational Colleges. *Frontiers in Educational Research*, *5*(14).

Kam, A. H., & Umar, I. N. (2023). Would gamification affect high and low achievers differently? A study on the moderating effects of academic achievement level. *Education and Information Technologies*, *28*(7), 8075-8095.

Kam, A. H., & Umar, I. N. (2024). Fostering autonomous motivation: a deeper evaluation of gamified learning. *Journal of Computing in Higher Education*, *36*(2), 368-388.

Kinio, A. E., Dufresne, L., Brandys, T., & Jetty, P. (2019). Break out of the classroom: the use of escape rooms as an alternative teaching strategy in surgical education. *Journal of surgical education*, *76*(1), 134-139.

Kirisci, N., Sak, U., & Karabacak, F. (2020). The effectiveness of the selective problem solving model on students' mathematical creativity: A Solomon four-group research. *Thinking Skills and Creativity*, *38*, 100719.

Koenig, S. T., Crucian, G. P., Dalrymple-Alford, J. C., & Dünser, A. (2009). Virtual reality rehabilitation of spatial abilities after brain damage. *Annual Review of Cybertherapy and Telemedicine 2009*, 105-107.

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International journal of information management*, *45*, 191-210.

Koivisto, M. (2022). Gamified learning of project business skills. *International Journal of Advanced Corporate Learning*, *15*(1), 45.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Kueh, Y. C., Abdullah, N., Kuan, G., Morris, T., & Naing, N. N. (2018). Testing measurement and factor structure invariance of the physical activity and leisure motivation scale for youth across gender. *Frontiers in Psychology*, *9*, 1096.

Lai, M., & Lo, L. N. (2006). Decentralization and social partnership: The development of vocational education at Shanghai and Shenzhen in China. *Educational Research for Policy and Practice*, *5*, 101-120.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lamlé, P., Keuch, L., Pukas, L., Reinsch, J., Röder, H., Clausner, H., & Watzke, S. (2023). Suicidal ideations among students attending vocational school in Germany: a cross-sectional study to identify prevalence and associated risk factors. *Journal of Public Health*, *31*(11), 1945-1952.

Landers, R. N. (2014). Developing a Theory of Gamified Learning:Linking Serious Games and Gamification of Learning. *Simulation & Gaming*, *45*(6), 752-768. https://doi.org/10.1177/1046878114563660

Latre‑Navarro, L., Quintas‑Hijós, A., & Sáez‑Bondía, M. J. (2024). The combined effects of an anatomy program integrating drawing and gamification on basic psychological needs satisfaction among sport sciences students: Results of a natural experiment. *Anatomical sciences education*, *17*(2), 366-378.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage publications.

Li, H., Lin, C. D., Bray, M. A., & Kehle, T. J. (2005). The measurement of stressful events in Chinese college students. *Psychology in the Schools*, *42*(3), 315-323.

Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review. *Educational technology research and development*, *72*(2), 765-796.

Li, X., Zou, L., & Li, H. (2024). Tai chi movement recognition and precise intervention for the elderly based on inertial measurement units and temporal convolutional neural networks. *Sensors*, *24*(13), 4208.

Lin, W. (2021). Innovative thinking in higher vocational colleges' physical education. *Journal of Contemporary Educational Research*, *5*(6), 14-18.

Lindberg, S. (2019). Gamification for self-directed learning in higher education. EDULEARN19 Proceedings,

Lovibond, S. H. (1995). Manual for the depression anxiety stress scales. *Sydney psychology foundation*.

Lu, Z., Mao, C., Tan, Y., Zhang, X., Li, Z. a., Zhang, L., Zhu, W., & Sun, Y. (2022). Trends in Physical Fitness and Nutritional Status among School-Aged Children and Adolescents during the COVID-19 Pandemic in Shaanxi, China—A Cross-Sectional Study. *Nutrients*, *14*(15), 3016.

Luarn, P., Chen, C.-C., & Chiu, Y.-P. (2023). Enhancing intrinsic learning motivation through gamification: a self-determination theory perspective. *The International Journal of Information and Learning Technology*, *40*(5), 413-424.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-386.

Mai, N. N., Takahashi, Y., & Oo, M. M. (2020). Testing the effectiveness of transfer interventions using Solomon four-group designs. *Education Sciences*, *10*(4), 92.

Maxwell, S. E., & Delaney, H. D. (2003). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (2nd ed.). Routledge. https://doi.org/10.4324/9781410609243

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research*, *51*(5), 339-343.

Meng, L. (2012). Investigation and Analysis on Students' Physical Health in Higher Vocational Colleges. *Journal of Inner Mongolia Normal University*.

Meng, L., & Ma, Q. (2015). Live as we choose: the role of autonomy support in facilitating intrinsic motivation. *International Journal of Psychophysiology*, *98*(3), 441-447.

Metrôlho, J., Ribeiro, F., Fonseca, E., Machás, L., & Dias, A. (2019). A mobile application for orienteering events in physical education subjects. ICERI2019 Proceedings,

Metwally, A. H. S., Tlili, A., Yousef, A. M. F., Huang, R., & Nacke, L. E. (2022). Empirical Evidence on Gamification and Learning Analytics (GaLA): What is Missing? 2022 International Conference on Advanced Learning Technologies (ICALT),

Mi, L., Zhang, W., Yu, H., Zhang, Y., Xu, T., & Qiao, L. (2024). Knowledge mapping analysis of pro-environmental behaviors: research hotspots, trends and frontiers. *Environment, Development and Sustainability*, 1-35.

Molanorouzi, K., Khoo, S., & Morris, T. (2014). Validating the physical activity and leisure motivation scale (PALMS). *BMC Public Health*, *14*(1), 909.

Moloney, A., & Maple, H. J. (2023). Developing, Choosing, and Using the Chemical Toolbox for Infectious Diseases Research. *ACS Infectious Diseases*, *9*(1), 2-4. https://doi.org/10.1021/acsinfecdis.2c00539

Montiel-Ruiz, F. J., Sánchez-Vera, M.-d.-M., & Solano-Fernández, I. M. (2023). Social networks and gamification in physical education: A case study. *Contemporary Educational Technology*, *15*(1).

Moreira, F., Durão, N., Pereira, C. S., & Ferreira, M. J. (2017). Mobile learning with gamification and augmented reality in Portuguese High Education. EDULEARN17 Proceedings, Moreno-Guerrero, A.-J., Alonso García, S., Ramos Navas-Parejo, M., Campos-Soto, M. N., & Gómez García, G. (2020). Augmented Reality as a Resource for Improving Learning in the Physical Education Classroom. *International Journal of Environmental Research and Public Health*, *17*(10), 3637. https://www.mdpi.com/1660-4601/17/10/3637

Morris, T., & Rogers, H. (2004). Measuring motives for physical activity. Sport and chance of life: Proceedings of 2004 international sport science congress,

Mujiono, M., & Gazali, N. (2021). Literature review: Physical education in the covid-19 pandemic. *JUARA: Jurnal Olahraga*, *6*(1), 50-63.

Nalyvaiko, O., Zhukova, O., Ivanenko, L., Shvedova, Y., & Nekrashevych, T. (2021). Gamification as a New Format of Projects Method in Blended Learning Conditions Studying Disciplines of the Pedagogical Cycle. *Revista Romaneasca Pentru Educatie Multidimensionala*, *13*(4), 17-30.

Nikulina, I., & Snezhkova, A. (2019). Study of the level of formation of motivation for learning activities of students. *Вестник Самарского университета. История, педагогика, филология*, *25*(3), 89-94.

Niu, Y. (2023). Integrated physical education and medicine in general physical education at universities in the age of educational technologies. *BMC medical education*, *23*(1), 466. https://doi.org/10.1186/s12909-023-04440-9

Norton, P. J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, stress, and coping*, *20*(3), 253-265.

Ntoumanis, N. (2001). A self‐determination approach to the understanding of motivation in physical education. *British journal of educational psychology*, *71*(2), 225-242.

Ntoumanis, N., Ng, J. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Lonsdale, C., & Williams, G. C. (2021). A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health psychology review*, *15*(2), 214-244.

Oei, T. P., Sawang, S., Goh, Y. W., & Mukhtar, F. (2013). Using the depression anxiety stress scale 21 (DASS-21) across cultures. *International Journal of Psychology*, *48*(6), 1018-1029.

Pampel, F. C., Krueger, P. M., & Denney, J. T. (2010). Socioeconomic disparities in health behaviors. *Annual review of sociology*, *36*(1), 349-370.

Pan, M., & Zhang, S. (2023). Visualization of Prediction Methods for Wildfire Modeling Using CiteSpace: A Bibliometric Analysis. *Atmosphere*, *14*(6), 1009. https://www.mdpi.com/2073-4433/14/6/1009

Pascu, M., & Soitu, L. . (2023). Constructivism, gamification, learning. . In *Studia Universitatis Moldaviae. Seria Ştiinţe ale Educaţiei*.

Pate, R. R. (1988). The Evolving Definition of Physical Fitness. *Quest*, *40*(3), 174-179. https://doi.org/10.1080/00336297.1988.10483898

Patrician, P. A. (2002). Multiple imputation for missing data. *Research in nursing & health*, *25*(1), 76-84.

Pedrelli, P., Nyer, M., Yeung, A., Zulauf, C., & Wilens, T. (2015). College students: mental health problems and treatment considerations. *Academic psychiatry*, *39*, 503-511.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, *29*(5), 489-497.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, *30*(4), 459-467.

Polit, D. F., & Gillespie, B. M. (2010). Intention-to-treat in randomized controlled trials: Recommendations for a total trial strategy. *Research in nursing & health*, *33*(4), 355-368.

Prontenko, K., Griban, G. P., Bloshchynskyi, I., Melnychuk, I., Popovych, D., Popovych, V., Yastremska, S., Dzenzeliuk, D., & Novitska, I. (2020). Improvement of students' morpho-functional development and health in the process of sport-oriented physical education. *Wiadomości Lekarskie*(73 (1)), 161-168.

Qiu, Y., Liao, K., Zou, Y., & Huang, G. (2022). A Bibliometric Analysis on Research

Regarding Residential Segregation and Health Based on CiteSpace. *International Journal of Environmental Research and Public Health*, *19*(16), 10069. https://www.mdpi.com/1660-4601/19/16/10069

Quan-chen, Z. (2013). Analysis of potential risk factors of chronic diseases among college students. *Chinese Journal of School Health.*

Rebar, A. L., Stanton, R., Geard, D., Short, C., Duncan, M. J., & Vandelanotte, C. (2015). A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. *Health psychology review*, *9*(3), 366-378. https://doi.org/10.1080/17437199.2015.1022901

Rieckmann, M. (2017). *Education for sustainable development goals: Learning objectives*. UNESCO publishing.

Riedl, A., & Schelten, A. (2013). Technical discussions as supportive interventions in the process of constructivist teaching and learning. In *From Diagnostics to Learning Success: Proceedings in Vocational Education and Training* (pp. 115-126). Springer.

ROMAN, A. F., & FELEA, M. I. (2022). THE MOTIVATION OF LEARNING IN STUDENTS. *Journal Plus Education*, *31*(2), 10-20.

Roychowdhury, D. (2012). *Examining reasons for participation in sport and exercise using the physical activity and leisure motivation scale (PALMS)* Victoria University].

Rubín, S. (2015). Gamification: From the communication and educational technology to critical and creative thinking. INTED2015 Proceedings,

Ryan, & Deci. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary educational psychology*, *25*(1), 54-67. https://doi.org/https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford publications.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology*, *61*, 101860.

Ryan, R. M., & Deci, E. L. (2024). Self-determination theory. In *Encyclopedia of quality of life and well-being research* (pp. 6229-6235). Springer.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational psychology review*, *32*(1), 77-112.

Salajegheh, M., & Hoseinyshavoun, A. (2018). the Role of motivation in learning. *Journal of Medical Education Development*, *13*, 172-174.

Salim, N. A., Tuffaha, M. G., & Brant, J. M. (2020). Impact of a pain management program on nurses' knowledge and attitude toward pain in United Arab Emirates: Experimental-four Solomon group design. *Applied Nursing Research*, *54*, 151314.

Samah, L. A., Ismail, A., & Hasan, M. K. (2022). The Effectiveness of Gamification for Students' Engagement in Technical and Vocational Education and Training. *International Journal of Advanced Computer Science and Applications*, *13*(9).

Şanlı, E., & Ersanli, K. (2021). The effects of psycho-education program based on logotherapy for the development of a healthy sense of identity in emerging adulthoods. *Current Psychology*, *40*(9), 4173-4188.

Schulz, K. F., Altman, D. G., & Moher, D. (2011). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, *154*(4), 291-292.

Scott, A. J., Webb, T. L., Martyn-St James, M., Rowse, G., & Weich, S. (2021). Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials. *Sleep medicine reviews*, *60*, 101556.

Seomun, G., Ban, S., & Park, J. (2022). Identifying the knowledge structure of electromagnetic fields and health research: Text network analysis and topic modeling. *PLoS One*, *17*(8), e0273005. https://doi.org/10.1371/journal.pone.0273005

Sharipovich, H. S. (2024). Pedagogical Aspects of Forming a Healthy Lifestyle in Students of Medical Higher Education in the Process of Physical Education. *International*

Journal of Formal Education, *3*(1), 108-118.

Shen, Y.-J., Xu, W.-Y., & Liu, R.-Q. (2021). Research on the Physical Fitness Test Results of Vocational College Students under Vocational Physical Education. 2021 IEEE 3rd International Conference on Computer Science and Educational Informatization (CSEI),

Shen, Y., Huang, L., & Wu, X. (2022). Visualization analysis on the research topic and hotspot of online learning by using CiteSpace—Based on the Web of Science core collection (2004–2022) [Original Research]. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1059858

Shephard, R. J. (1988). PAR-Q, Canadian Home Fitness Test and exercise screening alternatives. *Sports medicine*, *5*(3), 185-195.

Shi-gan, N. (2014). Prevalence of anxiety and its associated factors among vocational college students. *ournal of Taishan Medical College.*

Shou-d, Z. (2008). A Brief Discussion on Improving Vocational College Girl Students' Interest in Sports. *Journal of Wuxi Institute of Technology.*

Signori, G. G., Guimarães, J. C. F. D., Severo, E. A., & Rotta, C. (2018). Gamification as an innovative method in the processes of learning in higher education institutions. *International Journal of Innovation and Learning*, *24*(2), 115-137.

Silva, F., Ramos, J., & Analide, C. (2022). Applications of Virtual and Augmented Reality for Practical Application Learning with Gamification Elements. *IxD&A*, *53*, 191-212.

Simón-Chico, L., González-Peño, A., Hernández-Cuadrado, E., & Franco, E. (2023). The Impact of a Challenge-Based Learning Experience in Physical Education on Students' Motivation and Engagement. *Eur J Investig Health Psychol Educ*, *13*(4), 684-700. https://doi.org/10.3390/ejihpe13040052

Solomon, R. L. (1949). An extension of control group design. *Psychological bulletin*, *46*(2), 137.

Sotos-Martínez, V. J., Ferriz-Valero, A., García-Martínez, S., & Tortosa-Martínez, J. (2024). The effects of gamification on the motivation and basic psychological needs of secondary school physical education students. *Physical Education and Sport*

Pedagogy, *29*(2), 160-176.

Sriramatr, S., Berry, T. R., & Spence, J. C. (2014). An Internet-based intervention for promoting and maintaining physical activity: a randomized controlled trial. *American journal of health behavior*, *38*(3), 430-439.

Stanley, J. C., & Campbell, D. T. (1963). *Experimental and quasi-experimental designs for research*. Chicago: R. McNally.

State Council of the People's Republic of China. (2019). *Notice on issuing the national vocational education reform implementation plan* (Guo Fa [2019] No. 4). S. C. o. t. P. s. R. o. China. https://www.gov.cn/zhengce/content/2019-02/13/content_5365341.htm

Su, C.-H. (2016). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. *Multimedia Tools and Applications*, *75*, 10013-10036.

Su, C. (2016). The effects of students' learning anxiety and motivation on the learning achievement in the activity theory based gamified learning environment. *EURASIA Journal of mathematics, science and technology education*, *13*(5), 1229-1258.

Teixeira, P. J., Marques, M. M., Silva, M. N., Brunet, J., Duda, J. L., Haerens, L., La Guardia, J., Lindwall, M., Lonsdale, C., & Markland, D. (2020). A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. *Motivation science*, *6*(4), 438.

Tenório, K., Chalco Challco, G., Dermeval, D., Lemos, B., Nascimento, P., Santos, R., & Pedro da Silva, A. (2020). Helping teachers assist their students in gamified adaptive educational systems: Towards a gamification analytics tool. Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21,

Topîrceanu, A. (2017). Gamified learning: A role-playing approach to increase student in-class motivation. *Procedia Computer Science*, *112*, 41-50. https://doi.org/https://doi.org/10.1016/j.procs.2017.08.017

Van Bommel, M., Kwakman, K., & Boshuizen, H. P. (2012). Experiences of social work students with learning theoretical knowledge in constructivist higher vocational education: A qualitative exploration. *Journal of Vocational Education & Training*, *64*(4), 529-542.

van Lankveld, W., Linskens, F., & Stolwijk, N. (2021). Motivation for physical activity: Validation of the dutch version of the physical activity and leisure motivation scale (PALMS). *International Journal of Environmental Research and Public Health*, *18*(10), 5328.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of personality and social psychology*, *87*(2), 246.

Vasconcellos, D., Parker, P. D., Hilland, T., Cinelli, R., Owen, K. B., Kapsal, N., Lee, J., Antczak, D., Ntoumanis, N., & Ryan, R. M. (2020). Self-determination theory applied to physical education: A systematic review and meta-analysis. *Journal of educational psychology*, *112*(7), 1444.

Waldron, J. J., & Dieser, R. B. (2010). Perspectives of fitness and health in college men and women. *Journal of College Student Development*, *51*(1), 65-78.

Wang, D., Md Khambari, M. N., Wong, S. L., & Razali, A. B. (2021). Exploring Interest Formation in English Learning through XploreRAFE+: A Gamified AR Mobile App. *Sustainability*, *13*(22), 12792. https://www.mdpi.com/2071-1050/13/22/12792

Wang, F., & Boros, S. (2021). The effect of physical activity on sleep quality: a systematic review. *European journal of physiotherapy*, *23*(1), 11-18.

Wang, F., Wang, Y., & Hu, X. (2017). Gamification teaching reform for higher vocational education in China: A case study on layout and management of distribution center. *International Journal of Emerging Technologies in Learning (Online)*, *12*(9), 130.

Wang, H. (2014). Higher vocational college physical education situation analysis and reform. *Advanced Materials Research*, *1044*, 1611-1614.

Wang, J., Liu, Z., Hou, J., & Ge, M. (2024). Research Progress and Trend Analysis of Concrete 3D Printing Technology Based on CiteSpace. *Buildings*, *14*(4), 989.

https://www.mdpi.com/2075-5309/14/4/989

Wang, L., Song, Y., Manson, J. E., Pilz, S., März, W., Michaëlsson, K., Lundqvist, A., Jassal, S. K., Barrett-Connor, E., & Zhang, C. (2012). Circulating 25-hydroxy-vitamin D and risk of cardiovascular disease: a meta-analysis of prospective studies. *Circulation: Cardiovascular Quality and Outcomes*, *5*(6), 819-829.

Wang, W., & Lu, C. (2020). Visualization analysis of big data research based on Citespace. *Soft Computing*, *24*(11), 8173-8186.

Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement*, *71*(1), 231-244.

Winter, E., Costello, A., O'Brien, M., & Hickey, G. (2021). Teachers' use of technology and the impact of Covid-19. *Irish educational studies*, *40*(2), 235-246.

Wu, D., Yu, L., Yang, T., Cottrell, R., Peng, S., Guo, W., & Jiang, S. (2020). The impacts of uncertainty stress on mental disorders of Chinese college students: evidence from a nationwide study. *Frontiers in Psychology*, *11*, 243.

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41-49.

Xing, L. (2011). Analysis on Students' Learning Motivation in Higher Vocational School and the Suggestions. *Vocational and Technical Education*.

Xue, J., & Guo, Y. (2024). Exploring the Modular Teaching of Higher Vocational Physical Education Courses Oriented on Vocational Needs. *Contemporary Education and Teaching Research*, *7*(5), 235-240.

Yacef, K., Caillaud, C., & Galy, O. (2018). Supporting learning activities with wearable devices to develop life-long skills in a health education app. Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part II 19,

Yanli, X., & Danni, L. (2021). Prospect of Vocational Education under the Background of Digital Age: Analysis of European Union's "Digital Education Action Plan (2021-

2027)". 2021 International Conference on Internet, Education and Information Technology (IEIT),

Yi, C. (2017). Reformation Methods of Physical Education Teaching in Higher Vocational Colleges from the Perspective of New Media. https://doi.org/10.25236/eduer.2017.110.

Yi, X., Fu, Y., Burns, R. D., Bai, Y., & Zhang, P. (2019). Body mass index and physical fitness among Chinese adolescents from Shandong Province: a cross-sectional study. *BMC Public Health*, *19*(1), 81.

Yin, X., Wang, S., Wang, C., Wang, H., Zong, Z., & Ban, Z. (2024). Research Progress on Economic Forest Water Stress Based on Bibliometrics and Knowledge Graph. *Phyton-International Journal of Experimental Botany*, *93*(5), 843--858. http://www.techscience.com/phyton/v93n5/56743

Yu, H., Kulinna, P. H., & Lorenz, K. A. (2018). An Integration of Mobile Applications into Physical Education Programs. *Strategies*, *31*(3), 13-19. https://doi.org/10.1080/08924562.2018.1442275

Yu, H., & Mi, Y. (2023). Application model for innovative sports practice teaching in colleges using internet of things and artificial intelligence. *Electronics*, *12*(4), 874.

Zach, S., Bar-Eli, M., Morris, T., & Moore, M. (2012). Measuring motivation for physical activity: an exploratory study of PALMS-the physical activity and leisure motivation scale. *Athletic Insight*, *4*(2), 141-152.

Zainuddin, Z., Rasyidin, R., Zanzibar, Z., Aruni, F., & Nurmasyahyati, N. (2023). Andragogical principles in a gamification concept: how does it work for adult learners in an online class? *Journal of Applied Research in Higher Education*, *15*(5), 1632-1648.

Zarei, S., Memari, A.-H., Moshayedi, P., Mosayebi, F., Mansournia, M. A., Khoo, S., & Morris, T. (2016). Psychometric properties of physical activity and leisure motivation scale in Farsi: an international collaborative project on motivation for physical activity and leisure. *Archives of Iranian medicine*, *19*(10), 0-0.

Zeng, Y., Wang, G., Xie, C., Hu, X., & Reinhardt, J. D. (2019). Prevalence and correlates of depression, anxiety and symptoms of stress in vocational college nursing students from Sichuan, China: a cross-sectional study. *Psychology, health & medicine*, *24*(7), 798-811.

Zhan-ping, Z. (2006). Investigation into the Health Status and Sports Activities of Students in Vocational Colleges.

Zhang, C., Shi, L., Tian, T., Zhou, Z., Peng, X., Shen, Y., Li, Y., & Ou, J. (2022). Associations Between Academic Stress and Depressive Symptoms Mediated by Anxiety Symptoms and Hopelessness Among Chinese College Students. *Psychology Research and Behavior Management*, *15*(null), 547-556. https://doi.org/10.2147/PRBM.S353778

Zhang, L., & Chen, Y. (2021). Examining the effects of gamification on Chinese college students' foreign language anxiety: A preliminary exploration. Proceedings of the 2021 4th International Conference on Big Data and Education,

Zhang, X. (2022). Innovative Research on Smart Physical Education Teaching in Colleges and Universities Under the Background of Internet+. 2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022),

Zhang, Y. (2013). Research on physical education of higher vocational colleges. Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012): Volume 4,

Zhang, Y. (2024). Analysis of Innovative Talent Cultivation Strategy of Vocational Education under the Strategy of Developing the Country through Science and Education Based on Big Data Investigation. *Applied Mathematics and Nonlinear Sciences*.

Zhi-feng, F. (2012). Personality-based: Contemporary Higher Vocational Education Should not be Ignored Sports Angle. *Journal of Nanjing Institute of Physical Education*.

Zikas, P., Bachlitzanakis, V., Papaefthymiou, M., Kateros, S., Georgiou, S., Lydatakis, N., & Papagiannakis, G. (2016). Mixed reality serious games and gamification for smart education. European conference on games based learning,

Zin, Z., Abdullah, A., Ahmad, H., & Sahib, F. (2024). The constructivist learning theory: exploring key technological advancements in learning management systems.

International Journal of Modern Education (IJMOE), *6*(23), 585-597.

Zong-bing, M. (2018). Research on Programming Teaching Strategy Based on Gamified Learning in Primary School. *US-China Education Review*, *8*(9), 385-395.

APPENDICES

Appendix A

Demographic information form and PAR-Q

1.Sex: Male [ ] Female [ ]

2. Age: Under 18 [ ] 18 [ ] 19 [ ] 20 [ ] 21[ ] 22 [ ] 23 [ ] 24 [ ] 25 [ ] 26 [ ] Above 26 [ ]

3.What is your major

Nursing [ ] Rehabilitation [ ] pharmacy [ ]

The health benefits of regular physical activity are clear; more people should engage in physical activity every day of the week. Participating in physical activity is very safe for MOST people. This questionnaire will tell you whether it is necessary for you to seek further advice from your doctor OR a qualified exercise professional before becoming more physically active.

Please read the 7 questions below carefully and answer each one honestly: check YES or NO.	YES	NO
1) Has your doctor ever said that you have a heart condition OR high blood pressure?		
2) Do you feel pain in your chest at rest, during your daily activities of living, OR when you do physical activity?		
3) Do you lose balance because of dizziness OR have you lost consciousness in the last 12 months? Please answer NO if your dizziness was associated with over-breathing (including during vigorous exercise)		
4) Have you ever been diagnosed with another chronic medical condition (other than heart disease or high blood pressure)?		
5) Are you currently taking prescribed medications for a chronic medical condition?		
6) Do you currently have (or have had within the past 12 months) a bone, joint, or soft tissue (muscle, ligament, or tendon) problem that could be made worse by becoming more physically active?		
7) Has your doctor ever said that you should only do medically supervised physical activity?		

Appendix B-

1 Letter of Informed Consent (English)

2 Letter of Informed Consent (Chinese)

1 Letter of Informed Consent (English)

Study title: The impact of gamified learning in higher vocational college physical education classes on students' physical fitness, mental health, and learning motivation: A randomized controlled trial

Researcher Name: Yu Fang

Address: Nanchang Health Vocational and Technical College

Phone Number: 18679102428

Email Address: 262373510@qq.com

Supervisor Name: Sonthaya Sriramatr

Address: Faculty of Physical Education, Sports, and Health, Srinakharinwirot University

Phone Number: 0867792977

Email address: sonthase@g.swu.ac.th

Participant Name: ....

Phone Number: ....

Email Address: ....

I, participant, have read and understand the information letter for the study entitled "The impact of gamified learning in higher vocational college physical education classes on students' physical fitness, mental health, and learning motivation: A randomized controlled trial" and am giving my informed consent to participate in this study. I understand that:

The purpose of this study is to evaluate the effectiveness of gamified learning interventions on physical fitness, mental health, and learning motivation among higher vocational college students

I am agreeing to participate in completing this study

I will gain no privileges for participating in the study or suffer no penalties for not participating

My participation in this research is voluntary. I understand that I can fully withdraw from the study and have all collected data withdrawn from the database before data analysis begins

I can ask questions and refuse to answer questions at any time without consequences

I am aware that my name and other identifying information will be changed in an attempt to protect my identity. My information will only be seen by the researcher and supervisor

I am aware that this research will be used in a doctoral dissertation, and may be used in other publications, presentations, and future research

I understand that I need to complete the PAR-Q health screening questionnaire to ensure safe participation in physical activities

I understand that the study involves physical fitness testing, mental health assessments, and learning motivation surveys

I understand that the study duration is 16 weeks, including pre- and post-intervention assessments

I agree to participate in the study based on these terms.

Print Name: .... Sign Name: .... Date: ....

2 Letter of Informed Consent (Chinese)

研究题目：游戏化学习在高职院校体育课程中对学生体质健康、心理健康和学习动机影响的随机对照试验

研究者

姓名：方愉

单位：南昌健康职业技术学院

电话：18679102428

邮箱：262373510@qq.com

指导教师

姓名：Sonthaya Sriramatr

单位：诗纳卡琳威洛大学体育学院

电话：+660867792977

邮箱：sonthase@g.swu.ac.th

参与者

姓名：……………………………………………

电话：……………………………………………

邮箱：……………………………………………

本人已仔细阅读并理解了题为"游戏化学习在高职院校体育课程中对学生体质健康、心理健康和学习动机影响的随机

对照试验"的研究信息，现在给予知情同意参与本研究。本人理解：

本研究的目的是评估游戏化学习干预对高职学生体质健康、心理健康和学习动机的有效性

本人同意参与并完成本研究的全部内容

参与本研究不会获得任何特殊待遇，不参与也不会受到任何不利影响

本人的参与完全是自愿的。本人理解可以随时完全退出研究，并要求在数据分析前撤回所有已收集的数据

本人可以随时提出问题或拒绝回答问题，而不会产生任何不良后果

本人了解我的姓名和其他身份识别信息将被更改以保护我的身份。我的信息只会被研究者及其指导教师查看

本人了解本研究将用于博士学位论文，并可能用于其他出版物、学术报告和未来研究

本人了解需要完成 PAR-Q 健康筛查问卷，确保安全参与体育活动

本人了解研究期间将进行体质健康测试、心理健康评估和学习动机调查

本人了解研究持续时间为 16 周，包括干预前后的测试

本人基于上述条件同意参与本研究。

签名................................ 日期................................

Appendix C

Ethics Approval

Human Research Ethics Approval for Scientific Research Project

Project Title: The impact of gamified learning in higher vocational college physical education classes on students' physical fitness, mental health, and learning motivation: A randomized controlled trial

Applicant: Yu Fang

Application Number:NCHVC-2025RT-2506

Undertaking Institution:Nanchang Health Vocational And Technical College

Approval Comments:

1. After review, the submitted research project application for "The impact of gamified learning in higher vocational college physical education classes on students' physical fitness, mental health, and learning motivation: A randomized controlled trial" complies with the ethical principles and moral requirements for medical and experimental animal research. The applicant is approved to conduct the study as per the submitted research protocol.

2.During the research process, the study must adhere to the international Declaration of Helsinki, as well as China's Measures for Ethical Review of Biomedical Research Involving Humans (Trial Implementation) and ethical principles, moral standards, and related laws, regulations, practices, and systems relevant to Good Clinical Practice(GCP).No modifications to the research content are allowed without prior approval from this committee. If the research protocol or informed consent documents are modified, a written report must be submitted promptly to the Ethics Committee, and all ethical issues must be resubmitted for review by the Ethics Committee.

Nanchang Health Vocational And Technical College

January 2[illegible]25

Appendix D

Images of Indoor and Outdoor Physical Fitness Testing Sites

Figure D-1 Indoor 1-Minute Sit-Up Testing Site for Female Student

Figure D-2 Indoor Standing Long Jump Testing Site

Figure D-3 Outdoor 800-Meter Testing Site for Female Students

Figure D-4 Outdoor 50-Meter Testing Site

Appendix E

Depression Anxiety Stress SCALES-21 Items (DASS-21)

Please read each statement and circle a number 0, 1, 2 or 3 which indicates how much the statement applied to you over the past week. There are no right or wrong answers. Do not spend too much time on any statement.

The rating scale is as follows:

0 Did not apply to me at all

1 Applied to me to some degree, or some of the time

2 Applied to me to a considerable degree or a good part of time

3 Applied to me very much or most of the time

Over the past week	Did not apply to me at all	Applied to me to some degree, or some of the time	Applied to me to a considerable degree or a good part of time	Applied to me very much or most of the time
I found it hard to wind down	0	1	2	3
I was aware of dryness of my mouth	0	1	2	3
I couldn't seem to experience any positive feeling at all	0	1	2	3
I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)	0	1	2	3

I found it difficult to work up the initiative to do things	0	1	2	3
I tended to over-react to situations	0	1	2	3
I experienced trembling (e.g. in the hands)	0	1	2	3
I felt that I was using a lot of nervous energy	0	1	2	3
I was worried about situations in which I might panic and make a fool of myself	0	1	2	3
I felt that I had nothing to look forward to	0	1	2	3
I found myself getting agitated	0	1	2	3
I found it difficult to relax	0	1	2	3
I felt down-hearted and blue	0	1	2	3
I was intolerant of anything that kept me from getting on with what I was doing	0	1	2	3
I felt I was close to panic	0	1	2	3
I was unable to become enthusiastic about anything	0	1	2	3
I felt I wasn't worth much as a person	0	1	2	3
I felt that I was rather touchy	0	1	2	3
I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)	0	1	2	3

I felt scared without any good reason	0	1	2	3
I felt that life was meaningless	0	1	2	3

Appendix F

The Physical Activity and Leisure Motivation Scale (PALMS)

In responding to the following statements, think of the motives you have for the physical activity you do. Try not to spend time pondering over your responses. There are no right or wrong answers. Indicate how much your motives correspond with each of the statements. In each case 1 indicates strongly disagree and 5 indicates strongly agree.

I undertake physical activity	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
to earn a living	1	2	3	4	5
because it helps me relax	1	2	3	4	5
because it's interesting	1	2	3	4	5
because I enjoy spending time with others	1	2	3	4	5
to get better at an activity	1	2	3	4	5
because I perform better than others	1	2	3	4	5
because I get paid to do it	1	2	3	4	5
to do activity with others	1	2	3	4	5
to better cope with stress	1	2	3	4	5
because it helps maintain a healthy body	1	2	3	4	5
to define muscle, look better	1	2	3	4	5
be physically fit	1	2	3	4	5
because it makes me happy	1	2	3	4	5

to get away from pressures	1	2	3	4	5
to maintain physical health	1	2	3	4	5
to improve existing skills	1	2	3	4	5
to be best in the group	1	2	3	4	5
to manage medical condition	1	2	3	4	5
to do my personal best	1	2	3	4	5
to do something in common with friends	1	2	3	4	5
because people tell me I need to	1	2	3	4	5
because it acts as a stress release	1	2	3	4	5
to improve body shape	1	2	3	4	5
to obtain new skills/activities	1	2	3	4	5
because it's fun	1	2	3	4	5
because it was prescribed by doctor, physio	1	2	3	4	5
to work harder than others	1	2	3	4	5
because it keeps me healthy	1	2	3	4	5
to compete with others around me	1	2	3	4	5
to talk with friends exercising	1	2	3	4	5
to keep current skill level	1	2	3	4	5
to improve appearance	1	2	3	4	5

to improve cardiovascular fitness	1	2	3	4	5
because I enjoy exercising	1	2	3	4	5
to take mind off other things	1	2	3	4	5
to lose weight, look better	1	2	3	4	5
because I have a good time	1	2	3	4	5
to be with friends	1	2	3	4	5
to be fitter than others	1	2	3	4	5
to maintain trim, toned body	1	2	3	4	5

Appendix G

Design of Gamified Learning in Physical Education Classes

Week	Game name	Group	Game equipment	Objective	Elimination	Victory Condition	Game Duration	Scoring Method
1	Name Tag Tearing Game	Two groups, with 25 participants in each group	Each player has a name tag with their name attached to their back	Players must protect their own name tags from being torn off while attempting to tear off the name tags of the opposing team's members.	When a player’s name tag is torn off, the player is eliminated and must leave the playing area.	A team wins when all the name tags of the opposing team have been torn off	The game lasts for 30 minutes, with a 3-minute break after each round.	Participants who do not participate receive 0 points. Those who participate receive 1 point, and the winning team earns an additional 3 points.

2	Slalom Race	Students are divided into two teams with an equal number of participants in each team.	A track is set up on the field, with several cones placed at intervals along the course. Each team member is required to navigate around these markers in sequence.	Each player must navigate around all the markers on their assigned track and return to the starting point to hand the relay baton to the next teammate. Players are require	If a player fails to navigate around a marker or commits an infraction, they must return to the point of error and correctly navigate around the marker before continuing the race; otherwise, they are considered eliminated.	The team that successfully completes the slalom race with all members returning to the starting point first is declared the winner.	The game lasts for 30 minutes, with a 3-minute break after each round.	Teams that complete the race are awarded 1 point. The team that finishes first earns an additional 3 points. Teams that do not complete the race receive 0 points.

				d to maintain the correct route, passing around all markers without committing any infractions.				
3	Rules for the Multi-Person Three-Legged Race	Students are divided into two teams with an equal number of partici	The legs of two adjacent members in each team are tied together using ropes or cloth strips,	Team members must coordinate their movements to run quickly to the	If a team member falls or the ties become loose, the team must stop and fix the issue before continuing	The team that successfully completes the three-legged race and returns to the starting	Each round of the race lasts for 30 minutes, with a 3-minute break after the	Non-participants receive 0 points. Participants receive 1 point, and the

		pants in each team.	forming a single unit. All members of each team must move together as a group.	finish line and then return to the starting point. It is essential that team members maintain a synchronized pace to avoid falling or loosening the ties.	the race; otherwise, they are considered eliminated.	point first is declared the winner.	completion of the race.	team that finishes first earns an additional 3 points.
4	Campus Oriente	Students are divide	Several task points	Teams must use a	A team is considered	The team that complete	The game lasts for	Non-participants

	ering Game	d into 10 teams, with 5 members in each team.	are set up across the campus. Each task point includes a challenge or question that must be completed or answered through teamwork before the team can proceed.	map or follow clues to locate and reach all task points in sequence, successfully complete the challenges at each point, and finally return to the starting point. The team that	eliminated if it fails to complete a challenge at a task point within the allotted time or if a significant mistake is made.	s all task points and returns to the starting point first is declared the winner.	30 minutes.	receive 0 points. Participants receive 1 point, and the team that finishes first earns an additional 3 points.

				completes all tasks and returns to the starting point the fastest wins.				
5	Group Jump Rope Game Rules	Students are divided into 5 teams, each consisting of 10 members.	Students are divided into 5 teams, each consisting of 10 members.	The objective for each team is to have all members sequentially enter the rope and jump collectively as	A team is considered eliminated if it fails to achieve the minimum required number of jumps within the allotted time or if any member fails to enter the	The team that achieves the highest number of jumps within the allotted time without errors is declared the winner.	The game lasts for 30 minutes, with a 3-minute break between rounds.	Non-participants receive 0 points. Participants receive 1 point, and the team with the highest number of

				many times as possible without interruption. Teams must maintain synchronization to avoid stopping the rope or making mistakes.	rope and jump successfully.			jumps earns an additional 3 points.
6	Kangaroo Jump Game	Students are divided into 5 teams with	Each player stands inside a large burlap or cloth	Each player must progress along the	If a player falls or fails to pass through a marker in the	The team that successfully completes the jumping	Each round lasts for 30 minutes, with a 3-	Non-participants receive 0 points. Partici

		an equal number of participants in each team.	sack, holding onto the sack's opening, and prepares to jump forward.	course by jumping in the sack, passing through designated markers, and ultimately reaching the finish line. The goal for each team is to have all members comple	correct order while jumping, they must return to the point of error and restart the jump from there. Players who fail to complete the course are considered eliminated.	course with all members reaching the finish line first is declared the winner.	minute break between rounds.	pants receive 1 point, and the team that finishes the course first earns an additional 3 points.

				te the jumping course in sequence.				
7	Back-to-Back Ball Relay Game	Students are divided into 5 teams, each consisting of 10 members, with equal numbers in each team.	Each team is provided with one volleyball.	Two players from each team start at the starting line, holding a ball between their backs without using their hands. They must cooperate to	If the ball is dropped during the relay, the players must return to the spot where the ball fell and restart from there. Teams that fail to complete the relay are considered	The team that completes the relay with all members and returns to the starting point first is declared the winner.	The total game time is 30 minutes, with a 3-minute break after each round.	Non-participants receive 0 points. Participants receive 1 point, and the team that finishes the relay first earns an additional 3

				walk or run to the finish line, where they pass the ball to the next pair of teamm ates. The objecti ve is to keep the ball from falling and succes sfully comple te the relay.	eliminated .			points.
8	Frisbee Relay	Stude nts are	Each team is	Team memb	If the frisbee	The team that	The total	Non- partici

	Game	divided into 10 teams, each consisting of 5 members.	equipped with one frisbee to be used for the relay.	ers must sequentially pass the frisbee to the next player, ensuring that the frisbee remains airborne during the transfer and does not touch the ground. Each player	falls to the ground or the transfer fails, the relay must be restarted from the last successful catch. Teams that do not complete the entire relay are considered eliminated.	completes the frisbee relay with all members and returns the frisbee to the starting point first is declared the winner.	game time is 30 minutes, with a 2-minute break after each round.	pants receive 0 points. Participants receive 1 point, and the team that completes the relay first earns an additional 3 points.

				must catch the frisbee within the designated area, complete their throw, and then pass it to the next teammate. The team that successfully completes the relay with all memb				

				ers and returns the frisbee to the starting point first is declared the winner.				
9-16	The games from Weeks 1 through 8 are repeated from Week 9 through Week 16, following the same rules and objectives for each activity. This structure ensures that students have the opportunity to reinforce their skills and improve their performance through continued practice and competition.							

Appendix H

The evaluation standards for student physical fitness testing by the Ministry of Education of China (excerpt)

Table A-1 Physical Fitness Test Scoring Standards for Male University Freshmen and Sophomores

Grades	Individual Scores	Vital Capacity (ml)	50-Meter Dash (seconds)	Sit and Reach (cm)	Standing Long Jump (cm)	Pull-ups in One Minute (reps)	1000-Meter Run (min)
excellent	100	5040	6.7	24.9	273	19	3'17"
	95	4920	6.8	23.1	268	18	3'22"
	90	4800	6.9	21.3	263	17	3'27"
Good	85	4550	7.0	19.5	256	16	3'34"
	80	4300	7.1	17.7	248	15	3'42"
Pass	78	4180	7.3	16.3	244		3'47"
	76	4060	7.5	14.9	240	14	3'52"
	74	3940	7.7	13.5	236		3'57"
	72	3820	7.9	12.1	232	13	4'02"
	70	3700	8.1	10.7	228		4'07"
	68	3580	8.3	9.3	224	12	4'12"
	66	3460	8.5	7.9	220		4'17"
	64	3340	8.7	6.5	216	11	4'22"
	62	3220	8.9	5.1	212		4'27"
	60	3100	9.1	3.7	208	10	4'32"
Fali	50	2940	9.3	2.7	203	9	4'52"
	40	2780	9.5	1.7	198	8	5'12"
	30	2620	9.7	0.7	193	7	5'32"
	20	2460	9.9	-0.3	188	6	5'52"
	10	2300	10.1	-1.3	183	5	6'12"

Table A-2 Physical Fitness Test Scoring Standards for Female University Freshmen and Sophomores

Grades	Individual Scores	Vital Capacity (ml)	50-Meter Dash (seconds)	Sit and Reach (cm)	Standing Long Jump (cm)	Pull-ups in One Minute (reps)	1000-Meter Run (min)
excellent	100	3400	7.5	25.8	207	56	3'18"
	95	3350	7.6	24.0	201	54	3'24"
	90	3300	7.7	22.2	195	52	3'30"
Good	85	3150	8.0	20.6	188	49	3'37"
	80	3000	8.3	19.0	181	46	3'44"
Pass	78	2900	8.5	17.7	178	44	3'49"
	76	2800	8.7	16.4	175	42	3'54"
	74	2700	8.9	15.1	172	40	3'59"
	72	2600	9.1	13.8	169	38	4'04"
	70	2500	9.3	12.5	166	36	4'09"
	68	2400	9.5	11.2	163	34	4'14"
	66	2300	9.7	9.9	160	32	4'19"
	64	2200	9.9	8.6	157	30	4'24"
	62	2100	10.1	7.3	154	28	4'29"
	60	2000	10.3	6.0	151	26	4'34"
Fali	50	1960	10.5	5.2	146	24	4'44"
	40	1920	10.7	4.4	141	22	4'54"
	30	1880	10.9	3.6	136	20	5'04"
	20	1840	11.1	2.8	131	18	5'14"
	10	1800	11.3	2.0	126	16	5'24"

Table A-3 Body Mass Index (BMI) Scoring Chart for Male and Female University Freshmen and Sophomores (Unit: kg/m^2)

Grades	individual Scores	Male	Female
Normal	100	17.9~23.9	17.2~23.9
Underweight	80	≤17.8	≤17.1
Overweight		24.0~27.9	24.0~27.9
Obese	60	≥28.0	≥28.0

Table A-4 Evaluation Criteria and Weighting Scores for the "National Student Physical Fitness Standards" Test (University Students)

Evaluation Criteria	BMI	Vital Capacity	50-Meter Dash	Sit and reach	Standing Long Jump	Pull-ups in One Minute (Male)/ Sit-ups in One Minute (Female)	1000-Meter Run (Male)/ 800-Meter Run (Female)
Scores (100)	15	15	20	10	10	10	20

Appendix I

Physical Fitness Test Content Validity Evaluation Form

Dear Expert,

Thank you for participating in the content validity evaluation of the physical fitness testing instrument. To ensure the scientific rigor and validity of the test items, we kindly request you to rate each item based on its representativeness, relevance, and importance, and to provide your valuable feedback.

Rating Guidelines:

Representativeness: Does the test item effectively represent the dimension of physical fitness it aims to measure? (1 = Not representative at all, 5 = Highly representative)

Relevance: How relevant is the test item to the objectives of physical fitness assessment? (1 = Not relevant at all, 5 = Extremely relevant)

Importance: How important is the test item in the overall assessment of physical fitness? (1 = Not important at all, 5 = Extremely important)

Please provide a score between 1 and 5 in the appropriate columns based on your professional judgment. Additionally, you are encouraged to offer any suggestions or feedback in the final column, if applicable. Your ratings and feedback are crucial to optimizing the design of the physical fitness testing instrument.

Instructions:

Please conduct the evaluation as objectively and comprehensively as possible.

If any aspect of the evaluation is unclear, do not hesitate to contact us for further clarification.

Kindly complete and submit the evaluation form by the specified deadline.

Thank you once again for your support and cooperation.

Test Item	Representativeness (1-5)	Relevance (1-5)	Importance (1-5)	Suggestions
Height and Weight				
Vital Capacity				
Sit and reach				
Standing Long Jump				
Sit-ups (Female)				
Pull-ups (Male)				
50-meter Sprint				
800 meters (Female)				
1000 meters (Male)				

Appendix J

Chinese Version of the Questionnaire.

动感课堂：探索体育课游戏化学习对高职学生的影响

编号：__________

亲爱的同学：

欢迎参加我们的问卷调查！本次问卷旨在了解游戏化学习在高职院校体育课中的应用对学生身体健康、心理健康以及体育学习动机的影响。我们希望通过您的真实反馈，帮助我们更好地了解这一创新教学方法的效果，从而改进教学设计，为您和您的同学们创造更有趣、更有效的学习体验。您的回答没有对错之分，请放心填写，同时您的答案将被严格保密，仅用于研究目的。感谢您的参与与支持！

第一部分: 基本信息

1. 您的性别：

男[ ]　　女[ ]

2. 您的年龄：

18 岁以下[ ]　18 岁[ ]　19 岁[ ]　20 岁[ ]

21 岁[ ]　22 岁[ ]　23 岁[ ]　24 岁[ ]

25 岁[ ]　26 岁[ ]　26 岁以上[ ]

3. 您所在年级：

大一 [ ]　大二 [ ]　大三 [ ]

第二部分：抑郁焦虑压力量表 (DASS-21)

请阅读以下每个陈述，并在符合您自身情况那栏打"√"，以表示过去一周内该陈述在您身上的适用程度。答案无对错之分。请勿在任一陈述上花费过多时间。

在过去一周内：

项目	完全不适用	有时适用	较常适用	非常符合或大部分时间适用
1. 我感到难以放松	[ ]	[ ]	[ ]	[ ]
2. 我感到口干	[ ]	[ ]	[ ]	[ ]
3. 我完全无法感受到任何积极情绪	[ ]	[ ]	[ ]	[ ]
4. 我出现呼吸困难（如呼吸过快、非体力活动时的气短）	[ ]	[ ]	[ ]	[ ]
5. 我很难提起劲做事	[ ]	[ ]	[ ]	[ ]
6. 我对某些情况反应过度	[ ]	[ ]	[ ]	[ ]
7. 我感到身体颤抖（如手抖）	[ ]	[ ]	[ ]	[ ]
8. 我感到消耗了大量紧张能量	[ ]	[ ]	[ ]	[ ]
9. 我担心自己会在某些场合恐慌或出丑	[ ]	[ ]	[ ]	[ ]
10. 我感到对未来毫无期待	[ ]	[ ]	[ ]	[ ]
11. 我感到焦躁不安	[ ]	[ ]	[ ]	[ ]
12. 我感到难以放松	[ ]	[ ]	[ ]	[ ]
13. 我感到情绪低落	[ ]	[ ]	[ ]	[ ]
14. 我对任何阻碍我继续做事的事情感到不耐烦	[ ]	[ ]	[ ]	[ ]
15. 我感到濒临恐慌	[ ]	[ ]	[ ]	[ ]
16. 我对任何事情都提不起兴趣	[ ]	[ ]	[ ]	[ ]
17. 我感到自己毫无价值	[ ]	[ ]	[ ]	[ ]
18. 我感到自己容易生气	[ ]	[ ]	[ ]	[ ]

19. 我注意到心跳异常（如心跳加快、漏跳等，且非因体力活动引起）	[ ]	[ ]	[ ]	[ ]
20. 我无故感到害怕	[ ]	[ ]	[ ]	[ ]
21. 我感到生活毫无意义	[ ]	[ ]	[ ]	[ ]

第三部分：身体活动和休闲动机量表（PALMS）

在回答以下陈述时，请思考您参与体育活动背后的动机。无需花费过多时间思考答案，答案没有对错之分。请根据您的实际情况，选择与每项陈述的符合程度，并打"√"。

我进行身体活动是：

项目	非常不同意	不同意	不确定	同意	非常同意
1. 为了维持生计	[ ]	[ ]	[ ]	[ ]	[ ]
2. 因为它有助于我放松	[ ]	[ ]	[ ]	[ ]	[ ]
3. 因为它富有趣味	[ ]	[ ]	[ ]	[ ]	[ ]
4. 是因为我享受与他人共度时光	[ ]	[ ]	[ ]	[ ]	[ ]
5. 为了提高某项技能	[ ]	[ ]	[ ]	[ ]	[ ]
6. 因为我表现得比他人更好	[ ]	[ ]	[ ]	[ ]	[ ]
7. 因为我能因此获得报酬	[ ]	[ ]	[ ]	[ ]	[ ]
8. 为了与他人共同参与活动	[ ]	[ ]	[ ]	[ ]	[ ]

9. 为了更好地应对压力	[ ]	[ ]	[ ]	[ ]	[ ]
10. 因为它有助于维持健康的身体状态	[ ]	[ ]	[ ]	[ ]	[ ]
11. 为了塑造肌肉线条，提升外貌形象	[ ]	[ ]	[ ]	[ ]	[ ]
12. 为了保持身体健康	[ ]	[ ]	[ ]	[ ]	[ ]
13. 因为它让我快乐	[ ]	[ ]	[ ]	[ ]	[ ]
14. 为了远离压力	[ ]	[ ]	[ ]	[ ]	[ ]
15. 为了维持身体健康	[ ]	[ ]	[ ]	[ ]	[ ]
16. 为了提升现有技能	[ ]	[ ]	[ ]	[ ]	[ ]
17. 为了成为团队中的佼佼者	[ ]	[ ]	[ ]	[ ]	[ ]
18. 为了管理健康状况	[ ]	[ ]	[ ]	[ ]	[ ]
19. 为了发挥出个人最佳水平	[ ]	[ ]	[ ]	[ ]	[ ]
20. 为了与朋友有共同的活动	[ ]	[ ]	[ ]	[ ]	[ ]
21. 因为别人告诉我需要这样做	[ ]	[ ]	[ ]	[ ]	[ ]

22. 因为它可以缓解压力	[ ]	[ ]	[ ]	[ ]	[ ]
23. 为了塑造体形	[ ]	[ ]	[ ]	[ ]	[ ]
24. 为了掌握新技能或参与新活动	[ ]	[ ]	[ ]	[ ]	[ ]
25. 因为它很有趣	[ ]	[ ]	[ ]	[ ]	[ ]
26. 因为医生或理疗师的建议	[ ]	[ ]	[ ]	[ ]	[ ]
27. 为了比他人更加努力	[ ]	[ ]	[ ]	[ ]	[ ]
28. 因为它让我保持健康	[ ]	[ ]	[ ]	[ ]	[ ]
29. 为了与周围人竞争	[ ]	[ ]	[ ]	[ ]	[ ]
30. 为了与锻炼的朋友交流	[ ]	[ ]	[ ]	[ ]	[ ]
31. 为了保持现有技能水平	[ ]	[ ]	[ ]	[ ]	[ ]
32. 为了改善外在形象	[ ]	[ ]	[ ]	[ ]	[ ]
33. 为了提高心血管健康	[ ]	[ ]	[ ]	[ ]	[ ]
34. 因为我喜欢锻炼	[ ]	[ ]	[ ]	[ ]	[ ]

35. 为了转移注意力，摆脱其他烦恼	[ ]	[ ]	[ ]	[ ]	[ ]
36. 为了减肥，让自己看起来更棒	[ ]	[ ]	[ ]	[ ]	[ ]
37. 因为我享受其中乐趣	[ ]	[ ]	[ ]	[ ]	[ ]
38. 为了与朋友共度时光	[ ]	[ ]	[ ]	[ ]	[ ]
39. 为了比他人更健康	[ ]	[ ]	[ ]	[ ]	[ ]
40. 为了保持苗条、健美的身材	[ ]	[ ]	[ ]	[ ]	[ ]

感谢您的参与！祝您身体健康，学业有成！

Appendix K

Game Score Table

ID	Weeks	Total Score
1	[illegible]	[illegible]
2	[illegible]	[illegible]
3	[illegible]	[illegible]
4	[illegible]	[illegible]
5	[illegible]	[illegible]
6	[illegible]	[illegible]
7	[illegible]	[illegible]
8	[illegible]	[illegible]
9	[illegible]	[illegible]
10	[illegible]	[illegible]
11	[illegible]	[illegible]
12	[illegible]	[illegible]
13	[illegible]	[illegible]
14	[illegible]	[illegible]
15	[illegible]	[illegible]
16	[illegible]	[illegible]
17	[illegible]	[illegible]
18	[illegible]	[illegible]
19	[illegible]	[illegible]
20	[illegible]	[illegible]
21	[illegible]	[illegible]
22	[illegible]	[illegible]
23	[illegible]	[illegible]
24	[illegible]	[illegible]
25	[illegible]	[illegible]
26	[illegible]	[illegible]
27	[illegible]	[illegible]
28	[illegible]	[illegible]
29	[illegible]	[illegible]
30	[illegible]	[illegible]
31	[illegible]	[illegible]
32	[illegible]	[illegible]
33	[illegible]	[illegible]
34	[illegible]	[illegible]
35	[illegible]	[illegible]
36	[illegible]	[illegible]
37	[illegible]	[illegible]
38	[illegible]	[illegible]
39	[illegible]	[illegible]
40	[illegible]	[illegible]
41	[illegible]	[illegible]
42	[illegible]	[illegible]
43	[illegible]	[illegible]
44	[illegible]	[illegible]
45	[illegible]	[illegible]
46	[illegible]	[illegible]
47	[illegible]	[illegible]
48	[illegible]	[illegible]
49	[illegible]	[illegible]
50	[illegible]	[illegible]

ID	Weeks	Total Score
101	[illegible]	[illegible]
102	[illegible]	[illegible]
103	[illegible]	[illegible]
104	[illegible]	[illegible]
105	[illegible]	[illegible]
106	[illegible]	[illegible]
107	[illegible]	[illegible]
108	[illegible]	[illegible]
109	[illegible]	[illegible]
110	[illegible]	[illegible]
111	[illegible]	[illegible]
112	[illegible]	[illegible]
113	[illegible]	[illegible]
114	[illegible]	[illegible]
115	[illegible]	[illegible]
116	[illegible]	[illegible]
117	[illegible]	[illegible]
118	[illegible]	[illegible]
119	[illegible]	[illegible]
120	[illegible]	[illegible]
121	[illegible]	[illegible]
122	[illegible]	[illegible]
123	[illegible]	[illegible]
124	[illegible]	[illegible]
125	[illegible]	[illegible]
126	[illegible]	[illegible]
127	[illegible]	[illegible]
128	[illegible]	[illegible]
129	[illegible]	[illegible]
130	[illegible]	[illegible]
131	[illegible]	[illegible]
132	[illegible]	[illegible]
133	[illegible]	[illegible]
134	[illegible]	[illegible]
135	[illegible]	[illegible]
136	[illegible]	[illegible]
137	[illegible]	[illegible]
138	[illegible]	[illegible]
139	[illegible]	[illegible]
140	[illegible]	[illegible]
141	[illegible]	[illegible]
142	[illegible]	[illegible]
143	[illegible]	[illegible]
144	[illegible]	[illegible]
145	[illegible]	[illegible]
146	[illegible]	[illegible]
147	[illegible]	[illegible]
148	[illegible]	[illegible]
149	[illegible]	[illegible]
150	[illegible]	[illegible]

VITA