Chinese students' experiences of 'high-stakes' assessment: The role of fitness testing

Jing Yang[a], Gillian Bartle[b], David Kirk[ac] and Dillon Landi[c*]

[a]*Strathclyde Institute of Education, The University of Strathclyde, Glasgow, UK;*

[b]*School of Humanities, Social Sciences and Law, University of Dundee, Dundee, UK;*

[c]*School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, AUS;*

*Corresponding Author

d.landi@uq.edu.au

Chinese students' experiences of 'high-stakes' assessment: The role of fitness testing

Background: 'High-stakes' testing is a common practice in China and is expanding worldwide. As part of these practices, physical education assessments are compulsory for all students. Given this, China provides a unique setting to explore students' reflective experiences of mandatory 'high-stakes' assessment in physical education.

Purpose: The purpose of this paper was to explore Chinese students' reflective experiences and perceptions of the physical education component of the Zhongkao ('high-stakes' senior high school entrance examination).

Methods: This paper uses semi-structured group interviews with 24 students enrolled in a Year 11 physical education class in Shanghai. The transcripts from the interviews were analysed using concept coding, conceptual mapping, and analytical memos.

Findings: This paper found 'high-stakes' assessment in Shanghai focused on physical performances and students felt this went against the perceived holistic aims of physical education. The students also reported teachers changed their pedagogical approach to focus on achieving higher fitness scores. Lastly, students stated the test lacked diversity and was not a fair measurement of learning.

Conclusion: This study suggests 'high-stakes' testing practices in China are disconnected from the students' beliefs about the value of physical education. As such, we contend that assessment practices in physical education should reflect the diverse forms of learning that happens in physical education.

Keywords: physical education, fitness tests, assessment, China, Zhongkao

Word count: 8513

Introduction

Routine high-stakes assessment has become a dominant practice in junior high school physical education in China (Chen and Brown 2013). The term 'high-stakes' illustrates the centrality of these assessments because their results are used to inform significant educational decisions. In China, high-stakes assessments are used to evaluate teachers' (Supovitz 2009), motivate students (Gölogluu Demir and Kaplan Keles 2021), and inform admissions decisions to universities and secondary schools (Chen and Brown 2013). Thus, the use of high-stakes assessments is not a neutral tool that only measures student learning (Farvis and Hay 2020) but is a sociopolitical process that has consequences for people's lives (Ryan 2002).

There are two significant high-stakes assessments for students in China, the *Zhongkao* and Gaokao[1]. The first, the *Zhongkao,* is a Senior High School Entrance Exam (Wu 2015). The *Zhongkao* is taken by Junior High School students (~15 years old) and the results are used to inform decisions for admission to Senior High School, or Vocational High School (Chen and Brown 2013). The second test, the *Gaokao,* is a National Universities/ Colleges Entrance Exam (Jing and Liu 2019). The *Gaokao* is taken by Senior High School students (~18 years old) and the results are used to inform decisions for entry into Universities and Colleges. Both tests have immense impact on

[1] In this paper, we use the term 'Zhongkao' instead of the direct English translation of 'Senior High School Entrance Exam.' We also use 'Gaokao' instead of the direct English translation of 'National Universities/College Entrance Exam.' This is because the English translation of strips the cultural significance attached to these terms. The terms 'Zhongkao' and 'Gaokao' do not only mean assessment. Instead, they are societal 'buzzwords' (Gu, Ma and Teng 2017) that represent 'make or break' (Li 2023) cultural events where students that complete the exam process are even considered 'survivors' (Gu, Ma and Teng 2017).

students because the results determine the quality and prestige of schools they are able to attend (Chen and Brown 2013). As such, these exams place considerable pressure on schools, parents, teachers, and students (Wu 2015).

The Communist Party of China and State Council (CPC and State Council 2007) issued *Opinions on Strengthening Youth Physical Education and Enhancing Physical Fitness of Young People.* This document required each province to include physical education as part of the overall score in the *Zhongkao*. Importantly, each province (in some cases large cities) develops their own version of the *Zhongkao*. Given this, the exam may be different based on geographical location (João Pires 2019). Further, each province decides the 'weight' of the physical education component of the exam in relation to other subjects (e.g., Chinese Literacy, English, Maths) (Meng et al. 2021). Taking local context into account, the focus of this paper is on Shanghai *Zhongkao*.

Shanghai Zhongkao: Physical Education Component

The Shanghai *Zhongkao* is comprised of 7 subjects with a total of 750 points: Chinese Literacy (150 pts.), English (150 pts.), Ethics (60 pts.), History (60 pts.), Maths (150 pts.), Physical Education (30 pts.), and Science (150 pts.). Students can earn points for each subject in two ways: (a) standardised exam and (b) grades/marks given by teachers (e.g., report cards). The total number of points across all subjects accumulates into a final score used to determine where students are admitted into Senior High School. The higher the score, the better chance of admission into a prestigious school.

The physical education component of the Shanghai *Zhongkao* is worth 30 of the 750 overall points. There are two parts to the physical education component: internal evaluation (15 pts.) and external evaluation (15 pts.). The internal evaluation (15 pts) is

assessed by local physical education teachers and is based on yearly class grades. The focus of this paper is on the external evaluation (15 pts), which is a standardised test implemented across the entire school system (Shanghai Municipal Education Committee 2019). The physical education test has four categories assessing fitness and sport skills: (a) endurance (6 points), (b) muscular strength (3 points), (c) lifetime sport skills (3 points), and (d) team sport skills (3 points). External examiners administer the test annually during an exam period in April.

Despite only accounting for 30 of the 750 points, the physical education test has recently received greater attention amongst parents, students, and teachers. Further, the State Council has urged provinces to increase the point value of physical education to be equivalent with other subjects (Shuo 2020). This recommendation is based on government initiatives to improve public health outcomes and raise physical activity levels (Wang, Ha and Wen 2014). This is perhaps one reason why the physical education test does not assess all forms of student learning, but rather only physical performances. Given the 'high-stakes' nature of the *Zhongkao*, parents and students are taking physical education more seriously (Ni and Zhang 2021). Hu (2017) reported that some parents have even resorted to giving their children stimulants on exam days to improve performance. Thus, the high-stakes nature of the *Zhongkao* places considerable pressure on students to perform well on the physical education test.

Assessment in Physical Education: Privileging Fitness Tests

Generally, there have been three types of outcomes assessed in physical education: theoretical knowledge, motor skills, and fitness (López-Pastor et al. 2013). Theoretical knowledge usually includes topics like rules, tactics, cultural norms, and other cognitive

domain content. Motor skill assessments evaluate how well a student has mastered (or improved) specific sport or movement skills. Lastly, fitness tests attempt students' level of fitness or if fitness has improved over the duration of a unit. This is not to say other assessments do not exist, however, these three have been the most popular.

The debate about the usefulness of fitness testing in physical education has been polemical (Alfrey and Landi 2023). Some have argued fitness testing has benefits including increased student motivation under specific conditions (e.g., small groups, senior facilitators) amongst high-skilled students (Jaakkola et al. 2016; Simonton, Mercier and Garn 2019). Others claimed fitness testing can lead to educational benefits like teaching about goal setting and planning (McDonald and Trost 2015). There is also evidence of positive attitudes towards fitness testing amongst some students (Mercier and Silverman 2014; O'Keeffe, MacDonncha and Donnelly 2021). Proponents for fitness testing have maintained the test should be linked to educational aims (Keating 2003; Silverman, Keating and Phillips 2008) and this could *potentially* lead to a reduction in negative health outcomes (Keating et al. 2020). Supporters of fitness testing have rebutted criticisms of fitness testing by claiming 'the test is not the problem' but 'the *teaching approach* used for the test' is problematic (Keating and Silverman 2009, our emphasis).

On the other hand, fitness testing has many documented limitations. Within the previous literature, fitness testing was also linked to student a-motivation (Jaakkola et al. 2016) and low intrinsic motivation (Goudas, Biddle and Fox 1994). The studies above that showed positive attitudes toward fitness testing also documented statistically significant differences by gender – where boys overwhelmingly had influenced the

positive results (Mercier and Silverman 2014; O'Keeffe, MacDonncha and Donnelly 2021). When considering diversity, fitness testing has produced negative experiences for girls (Wrench and Garrett 2008) as well as Black, Latina/o, and LGBTQ+ students (Safron and Landi 2022). Further research has documented students more broadly do *not* enjoy fitness testing in physical education (Alfrey and Gard 2014; Hopple and Graham 1995). These results led Landi (2023) to ask:

> A large question looms over the fitness testing debate: does the implementation of fitness tests in physical education place the *potential* benefits of the assessment above the *empirically documented* negative affects it has on (diverse) young people? (Landi 2023, 5, original emphasis)

Other issues raised about fitness testing included: (a) questionable validity and reliability (Cale and Harris 2009); (b) lack of resources to conduct testing appropriately (Alfrey and Landi 2023); (c) ethical concerns around data privacy (Pluim and Gard 2018); and (d) cultural relevance and narrow perspectives of 'health' (Safron and Landi 2022). Further, policy research has documented there is no evidence that the use of fitness tests in physical education can improve health outcomes (Landi, Walton-Fisette and Sutherland 2021). Despite these acknowledged concerns and ongoing debates, there is limited research exploring students' experiences with high-stakes assessments.

Juxtaposition between Chinese culture and physical education

Another important aspect to consider is Chinese cultural traditions. Traditional Chinese perspectives of health are historically rooted in practices that have evolved and progressed over thousands of years. Traditional Chinese health takes a holistic and multi-dimensional view of the body that integrates the mind, body, and spirit to

maintain balance (Sun et al. 2013). With relation to human movement, Qigong is a key concept of Chinese health (Chen et al. 2019) that uses dynamic movements, breathing techniques, and mental focus to balance the body (Jiuzhang and Lei 2009). Importantly, the body is interconnected with internal and external factors that cut across emotional, spiritual, environmental, and social dimensions (Sun et al 2013).

School-based physical education in China, however, has been shaped by Western perspectives of health that 'train' the (isolated) physical body (Jin 2013). Tsai and Zhou (2016) traced the history of Chinese physical education showing roots in gymnastics (1890-1911), with a shift to military preparation (1912-1949), and then a dominance of sport skills (1949 to present). This history reflects trends in Western physical education more broadly (Kirk, 2020). The 2000s, however, brought a curricular shift with the inclusion of health into physical education (Jin 2013) resulting in a shift away from behaviourist approaches of skill learning to constructivist student centred philosophies (Wang, Ha and Wen 2014). Assessment practices, however, still reflect the behaviourist tradition that measures physical outcomes (Whittle et al. 2017).

The juxtaposition between Chinese culture and physical education is important. For example, Chinese physical educators often adopt teaching perspectives linked to apprenticeship and nurturing practices (Wang, Ha and Wen 2014). Yet, historical teaching and learning practices have been dominated by an instrumentalist view of training the body. This difference between beliefs and curriculum goals (Wang, Ha and Wen 2014) is just one reason why the juxtaposition between Chinese culture and physical education practices is important. Our interest is how this comes to matter for students' and their experiences of high-stakes assessment in physical education.

Pragmatic Paradigm, Purpose and Research Questions

This paper is situated within a pragmatic paradigm of research (Leavy, 2017). A main purpose of pragmatic research is to ground design and methods in a values-based approach that focuses on lived experiences (Landi, 2024) to address social problems (Kuhn, 1962). In so doing, pragmatic research does not adopt a specific theory or model (e.g., feminism, socioecological model) to frame the study but instead uses prior insights and research practices from their research community (Morgan, 2014). There is no doubt that fitness testing is considered one of the most debated physical education practices in the field (Alfrey & Landi 2023). The use of high-stakes testing in China is also highly controversial (Chen and Brown 2013). Our research was informed by the academic literature in high-stakes assessments and fitness testing in physical education.

China is a unique place to study fitness testing for several reasons. One, there is little research (in Anglophone literature) on Chinese students' experiences of fitness assessments. Further, testing in China is mandatory and high-stakes which has serious consequences. Additionally, there is a difference in Chinese perspectives of health and fitness testing practices. Given these considerations, the purpose of the paper was to explore Chinese students' reflective experiences and perceptions of the physical education component of the Shanghai *Zhongkao*. The research questions that informed the study were:

(1) What are the reflective experiences of Chinese students' engagement in the physical education component of the Shanghai *Zhongkao*?

(2) How do Chinese students perceive the usefulness of the physical education test in the Shanghai *Zhongkao*?

Research Methods and Design

The design of this research was a qualitative exploratory study using conversational inquiry (Leavy 2017). It was important for this work to be exploratory (Creswell 2014) because of the minimal research on Chinese students' experiences in fitness testing. A conversational approach was adopted because it is best suited to understand reflective experiences and the meaning created because of them.

Setting and Participants

This study took place at Sheng Wang School (pseudonym), a prestigious private secondary school in Shanghai, China. Sheng Wang School is a co-educational school that combines Junior and Senior levels. Sheng Wang School was conveniently selected (Marshall & Rossman 2006) because Jing (pseudonym, first author) was a teacher at the school for five years and used professional relationships to enter the research site. The participants were 24 students between 16-17 years old enrolled in a Year 11 physical education class. Each student participated in the *Zhongkao* assessment two years prior at different schools. Given this, students' accounts of experiences in testing were reflective (Kvale 1996) in nature.

Data Generation

Data were generated using four semi-structured group interviews (Kvale 1996) that comprised of six students each and lasted 40-50 minutes each. Jing and Gillian (second author), chose group interviews based on the *reflective* design. By reflective, we mean the interviews are meant to be about participants' experiences, knowledge, and viewpoints on a particular topic (Kvale 1996). This is different than getting a 'truthful'

or 'accurate' account of a phenomenon (Randall and Phoenix, 2009) which aims to state what happened with as little bias as possible. Put differently, our focus was on how students made meaning of their previous experiences (Kvale 1996).

Group interviews (Marshall and Rossman 2006) were useful here because they produce dialectic encounters (Rubin and Rubin 2005) where knowledge is constructed through engagement with multiple viewpoints. Therefore, young people generated knowledge about their experiences engaging in the Shanghai *Zhongkao* through conversations. Yet, this also led to questioning, challenging, and/or providing alternative experiences. All interviews were conducted via Teams, audio-recorded (in Mandarin) and transcribed verbatim (into written Chinese). They were translated from Chinese into English by Jing and reviewed by a second person with translation experience to ensure accuracy. All participants were given pseudonyms.

Data Analysis

Data analysis was an iterative process with multiple readings, coding and tagging, writing up, as well as revising. Jing first became familiar with the data by listening to audio-recordings, transcribing, and translating. Jing and Dillon (fourth author) sat with the transcripts and read them separately. Afterwards, they jointly used 'concept coding' (Saldaña 2013), where they read transcripts and tagged specific excerpts with a code (e.g., 'dislikes fitness testing') that represented broader meaning. Jing and Dillon then mapped the codes (Ringrose and Coleman 2013). Here, Jing and Dillon placed the tagged excerpts on sticky notes, compared them to each other, and re-constructed broader 'categories' into a visual map. The map changed through discussions where concepts were questioned, refined, and further developed. We settled on five 'insights'

(e.g., valuing PE, not enjoying tests, teachers changing pedagogy, tests are important, tests should be diverse).

Jing and Dillon then co-wrote analytical memos (Marshall and Rossman 2006) on each 'insight' and sent them to Gillian and David (the third author), who provided critical feedback. Using feedback, Dillon and Jing revised the five 'insights' and collapsed them into three: (a) students value holistic PE; (b) students dislike test and teaching approach change; and (c) test needs updating. These themes were sent back to Gillian and David for additional feedback. This led to a revision of findings: (a) Test not aligned to goals of physical education; (b) Pedagogy changed for (fitness) test; and (c) Assessment lacks diversity. After peer review, we integrated the suggestions from reviewers, and had multiple discussions and revisions. This led to three finalised results: (a) Testing defeated the perceived value of physical education; (b) Teaching approach changed for (fitness) testing; and (c) The test lacks inclusivity of diverse activities and bodies.

Ethics and Rigour

Like Smith and McGannon (2018), we prefer to use the term 'rigour' instead of 'trustworthiness'. This is because trustworthiness is part of historical residue that judges qualitative research using fixed criteria (Smith, Sparkes and Caddick 2014) based on measures connected to quantitative perspectives of research (Smith and Pheonix 2019). Instead, we believe high-quality research follows a 'rigorous' process that includes: (a) transparency (Strom and Martin 2017), (b) reflexivity (Alvesson and Sköldberg 2000), and (c) acknowledgement/ awareness (Koro-Ljungberg et al 2009).

For transparency, this study was Jing's master's dissertation at Dundee University where it was approved by the ethics committee. We intentionally went to great lengths to detail the step-by-step process from design to write-up that illustrated the rigorous process we went through. We purposefully omitted traditional validation methods (i.e. member checking, triangulation) because of epistemological differences (Landi 2024). Yet, we hope our detailed explanation of the process (Strom and Martin 2017) illustrates our ethical commitments to transparency and rigour.

Reflexivity is important to qualitative inquiry because it states who we are and how we influenced the research process (Alvesson and Sköldberg 2000). Jing is originally from China and was a physical educator for five years at Sheng Wang School (2010-2015) before being a postgraduate researcher in the UK. She used her professional network to recruit the school and participants. Her experiences as a former teacher in the school gave her insider knowledge to frame the study with relevance to the school and Chinese culture. Thus, Jing brings insight and cultural knowledge and all those biases, and benefits, that comes with them.

Gillian is a Lecturer and former physical educator in the UK. She is completing her PhD with a focus on young people's embodied experiences in physical literacy. As Jing's master's supervisor, she shaped the project by maintaining a focus on students' lived experiences during testing. David and Dillon are Jing's PhD supervisors. David is a full professor and veteran scholar in physical education. With his extensive knowledge of curriculum, history, and pedagogy in physical education, he challenged the team to engage deeply with literature and make connections to prior insights. Dillon is originally from the USA but has worked in universities across the world. He has an

interest in health and physical education as well as advanced qualitative data analysis using contemporary methods. His work in this area influenced the data analysis process by working with Jing every step of the way.

As a reflexive point, each step of this process was affected by the authors in different ways. This is due to our diversity in upbringing, cultures, genders, ages, interests, and theoretical traditions. Yet, we *acknowledge* and are *aware* this paper has limitations based on its epistemological positioning (Koro-Ljungberg et al 2009). The students are drawing on memories from two years previously and may forget or even mis-remember events. Further, three of us are not Chinese and it took extensive time to get a basic understanding of the cultural and curriculum context. Not to mention, none of us are young anymore and cannot truly understand students' lived experiences. Thus, this paper is a (re-)presentation of data in a way we felt offered insight to the field and ethically aligned to (re-)present these young people's experiences.

Findings

Below, we outlined three findings: (a) Testing defeated the perceived value of physical education; (b) Teaching approach changed for (fitness) testing; and (c) The test lacks inclusivity of diverse activities and bodies.

Testing defeated the perceived value of physical education

Most students had positive views about the value of physical education in schools. Unlike other subjects, the students believed physical education was able to address the range of diverse learning areas (not just intellectual) and aided in holistic development. For example, Wen Li (male) said:

> After all, students should develop morally, intellectually, physically, aesthetically, and comprehensively. Physical education can do this and is a very important subject.

Wen Li's statement about physical education was reflective of students' holistic beliefs regarding health and the body. To these students, physical education played a 'very important' role in developing the whole body that included physical, moral, intellectual, and other dimensions. The students stated that physical education was unique because it provided holistic learning experiences. Zhan Ping (male) stated:

> I think the ultimate goal of physical education was to have fun. For example, if you play and move, it is your *whole* body. If we just sit in the classroom and do mental work, it doesn't include all the things that are involved in learning. It is more complex than that.

Zhan Ping makes three important points. Firstly, Zhan stated physical education has an 'ultimate goal' of addressing the enjoyment component of the affective domain (fun). Many of the students agreed with this sentiment and stated that they believed physical education should prioritise experiencing pleasure in and through movement. This did not mean that physical education was *only* about enjoyment. This is because Zhang's stated learning was not just a 'mental' process but involved the 'whole' body (the second point worth reflecting on). Therefore, addressing the affective domain is important for these young people because it is a pre-requisite to engage in the multi-dimensional aims of holistic learning in movement. Thus, Zhan's final point was crucial: learning is complex and physical education is unique because it can address this complexity. These students valued physical education as not just 'movement' or

'enjoyment' but rather it was a holistic learning experience mediated through the body, self, and its environment.

The students were adamant about the importance of holistic development of the body (e.g., physical, cognitive, social) in physical education. In particular, the affective domain was a dimension that kept coming up as crucial. There was a belief that physical education should provide enjoyable experiences because an important goal was to develop positive feelings toward movement. Fan Yahui (female) stated:

> I think the essence of physical education is to strengthen the body and, most importantly, to develop students' interests in different forms of physical activity.

Fan Yahui stated that physical education plays a role in developing the physical body (strengthening) as well as positive attitudes (interests) in order to engage in movement. Many of these statements, show how the students believed physical education is meant to be multi-dimensional and holistic experience as a way to foster interest in engaging in different forms of human movement.

Despite students valuing the holistic nature of physical education, they believed high-stakes assessments went against this perspective of health. So, whilst the students believed the body and health should be considered holistic, the performance-based physical assessments contradicted their beliefs. Xian Huo (female) stated:

> The test was not about health. It was about achieving higher grades. Some people may not be able to achieve these grades and instead do damage to their bodies. I feel it's a bit contradictory to the purpose of physical education. Not to mention, it makes students feel bad and exhausted both physically and mentally.

Xian Huo raised several concerns about the high-stakes test. She stated the nature of the *Zhongkao* meant learning was not holistic but rather driven by performance scores. For Xian Huo, the test is *not* about health or learning, but only about physical ability and exertion. When students participated above their physical capacity, they did 'damage' to their bodies. Not only does this foster negative experience but also went against what is considered healthy. Thus, the emphasis on performance-based testing, gave students negative experiences that went against their core beliefs about health and the purpose of physical education.

The students were also critical of the test because it only focused on physical dimensions of the body and did not assess students in other learning areas of class. Tang Zhen (male) said:

> There was not much understanding of knowledge because the physical education test does not have a theoretical examination. It was too performance driven and defeated the original purpose of physical education and the *Zhongkao.*

Tang Zhen criticised the test because it only focused on physical performances whereas he believed, with most students, there are other dimensions of knowledge to assess. By limiting the assessment to physical performances, Tang Zhen believed the *Zhongkao* went against the broader aim of physical education (to develop the body holistically) because it became unidimensional.

One of the reasons why students felt the test defeated the purpose of physical education is because how it made them feel when taking it and even two years later. Below are examples:

> Every time I trained for the test in physical education, I was nervous. I was even more nervous when I took the physical education test. My hands were shaking and I was very nervous. (Zou Shui, female)
>
> For students who are not so good at sports, this test is a burden. For me, the thought of running is still a mental burden. In fact, the test did not inspire me to participate in physical activity. If anything, it did the opposite. (Tian Lin, male)
>
> The presence of a scoring system led students to train and overexert themselves for the sake of getting a high score. Many students end up doing a lot of damage to their bodies because they believe the pain and injury is worth the cost. (Li Huo, male)
>
> The positive side of the test is I got the marks I needed to get into a good high school. The negative side is that I never plan on running ever again. (Mao Yi, female)

Throughout the interviews it was evident that students held negative emotions and feelings about the test. The test brought about anxiety during movement and made the students feel discouraged to engage in movement. Many students stated they witnessed injuries, pain, and physical damage to their own and friends' bodies. Yet, these students still participated because of the implications the test had on high school admissions. This combination of experiencing pressure, anxiety, and in some cases pain, brought about negative feelings and emotions of physical education.

Teaching approach changed for (fitness) testing

A key finding was that the students stated their teachers changed their teaching approach to prepare for the physical fitness portion of the test. Lei Su (female) stated:

> The preparation for the physical education test is a very sudden feeling. In the first two years of physical education, everyone was quite happy and enjoyed physical education. There were no special requirements for running and students could develop at their own pace. But in the third year (Year 9), it felt as if all these test items were coming at you all at once and you had to do them very hard to achieve a good mark.

Lei Su's remarks provide insight into pedagogical changes in response to the high-stakes test administration in their third year. Lei Su stated that students enjoyed physical education in their first two years of junior high and enjoyed the focus on individual development. In the third year, however, the teachers changed their approach to 'teach to the test' so students would achieve a good mark. Importantly, the students stated that teachers *only* focused on training for fitness testing items and not sport skills. One way to do this, according to Lei Su, was to provide students specific target scores in different exercises so they can achieve a high mark to increase their *Zhongkao* score Here, physical education was no longer about learning, enjoyment, or development. Rather, it became fitness training to achieve scores.

An interesting point that students reported was the fitness training sessions were held during physical education class as well as other times. Due to an emphasis on achieving high scores, physical educators provided 'fitness sessions' outside of allocated class time. Gu Ming (male) explained:

> Intensive training starts two months before the physical education test. The teachers organised us to come to school in the morning and run and do morning exercises before class. Then they gave you additional items you can train for on your own later.

Many students expressed similar experiences to Gu Ming. That is, physical educators were focused on improving fitness scores, so they changed their pedagogy 2-3 months preceding the test. These new activities included training for the test items on the fitness section with a focus on strength, endurance, and speed training. Further, teachers organised before and after school sessions to do endurance training. These fitness sessions were not designed around student learning or development. Rather, they were 'exercise sessions' aimed at achieving higher scores.

Another point to highlight is that all students stated the training was centred on fitness. Despite the test including sport skills, teachers made the decision to focus solely on the fitness section. This is interesting because it was not *any* test that changed teachers' teaching approach. Rather, it was the *fitness test* that the students were training for. Yet, the pedagogical approaches used (as well as the test itself) had negative effects on student experiences. With very few exceptions, most students said they disliked physical education in the period directly preceding the *Zhongkao*. They also expressed negative feelings about participating in the *Zhongkao*. Here are some statements made about student experiences:

> Two or three months before the physical education test, class turned into intensive training. Personally, I think it's a bit of torture because there is no physical fitness training up until then. Just normal physical education where we learn and play different games. (Song Su, male)

> Physical education class directly before the physical education test was boring and painful. I really don't think this kind of physical education class was my cup of tea. (Yu Hu, female)

Overall, students expressed a general dissatisfaction when teachers changed their pedagogy to fitness preparation for the *Zhongkao*. The students stated strong words about their experiences including 'torture', 'painful', 'boring', and overall disinterest. This was a change in tone from when students expressed their beliefs around the value of physical education. Thus, the changes that teachers made in pedagogy led to students disliking physical education. In this case, scoring high on the fitness test was more important than the students' feelings during physical education.

The test lacks inclusivity of diverse activities and bodies

Students were adamant that the current test is not diverse enough and fails to meet the needs of all students. They stated the test needs to change to reflect the diversity of students' bodies and forms of human movement. Some students stated:

> I feel the physical education test has very few sports items in their options. This results in some students' skills not being fully explored. This is because everyone's strengths are different because everyone participates in different sports. Some of the students' sport skills may be overlooked. (Kong Fu, male)

> There could be more items on the test, and this would give students the option to pick their favourite form of physical activity rather than telling them what sports they have to participate in. (Wen Jie, female)

The above comments about the *Zhongkao* were echoed by many students. Notably, the test has two categories of sport skills: 'lifetime sport' (table tennis, badminton, tennis, martial arts or gymnastics) and 'team sport' (football, basketball or volleyball). The students must choose one topic from each category. Despite having this choice, the students were disappointed in the range of choices and argued they participated in a

range of diverse activities. The lack of options made students feel disconnected to the test and felt they could score better if it were more inclusive.

Another issue the students identified was the fitness portion of the test is unfair and does not consider individual differences. Further, the test assesses students by the same exercises, but the students felt there are a diverse range of bodies and there are multiple ways to be physically fit. Two students stated the following:

> The physical education test is unfair. It does not take into account our differences or the physical condition we were in before we started training to now. Also, there are other differences that aren't included like some girls might get their period that day when they have to take the exam. It's very unfair. (Gao Fen, female)
>
> It is not reasonable to include things in the exam like height and weight on your score. Overweight students like me, it discriminates against my size because I just have a different body type. (Mo Zhou, male)

Young people have different lives, diverse bodies, and a range of interests and experiences. The fitness portion of the exam, however, treated all students the same and awarded points based on performances in specific exercises. As Gao Fen stated, the test does not consider where a student started and perhaps improved their fitness. Rather, the test focused on outcomes regardless of students' circumstances. Further, the test is held once annually. This placed tremendous pressure on everyone but especially students who may not be able to perform well under pressure, who may have an injury, or different needs (e.g., menstruation). Another issue is that the test measures height and weight and includes this as part of the score. Here, the body itself is being judged not even for performance but instead uses flawed indicators that do not measure learning. Thus, students are not just judged on their performance but also on their bodies.

Therefore, the students believed the test was too narrow in focus, failed to address diversity, and was based on unfair measures. The students were also quick to highlight that their peers had a range of abilities that the test fails to capture. Thus, the students stated the test needs to change to be reflective of the needs of young people.

Discussion

Like research internationally, the Chinese students in this study maintained a positive perspective of physical education (Silverman 2017). The students valued physical education based on the belief that the subject is uniquely positioned to address the multidimensional and holistic aspects of health, the body, and learning (Quennerstedt, Landi and Casey 2024). This holistic view of health aligned with Chinese cultural beliefs that approach the body as balanced and interconnected (Sun et al. 2013). Thus, there is an ostensible synergy between Chinese culture and students' perspectives of the value of physical education. The *Zhongkao*, however, focused only on physical performance and reduced the complexity of health and the body to a number based on a single test. This reduction defeated the holistic nature of health and the body (Jiuzhang and Lei 2009) and the main reason these students came to value physical education.

The focus on physical performance in the *Zhongkao* presented a conflict of priorities between Chinese culture, the curriculum, physical education practices, and the assessment. The students overwhelmingly stated an important aspect of physical education being holistic was addressing affective outcomes, those components related to feelings and emotions (e.g., attitude, motivation, enjoyment) (Kirk 2020), to foster positive engagement in movement. They believed these outcomes were needed in the

school day to release stress but also a pre-requisite to develop the body holistically and show the multi-dimensional aspect of learning. By only focusing on physical performances, the test was not relevant to the holistic outcomes students valued.

Notably, students reported they enjoyed physical education during non-testing years. This provided an indication that teachers may know how to structure lessons in ways that prioritise student learning across all five domains of learning (Quennerstedt, Landi, and Casey 2024). It was also in line with previous research that indicated Chinese teachers emphasize apprenticeship and a nurturing culture (Wang, Ha and Wen, 2014). Directly preceding the test, however, students in our study reported drastic changes in teaching approach where the aim was to score well on the *fitness test*. Notably, this did not include the sport skills portion of the test.

It has been previously argued that fitness tests are neutral tools and the teaching approach to fitness has been problematic (e.g., Keating and Silverman 2009). Our evidence, on the other hand, suggested that teachers changed their teaching practices in response to preparing for the test. As such, fitness tests were not a neutral tool but may be part of broader power relations around high-stakes testing that places pressure on teachers to 'teach to the test' (Wu 2015). From this perspective, assessments and teaching are interconnected and reciprocally affect one another in relation to student learning (Quennerstedt 2019). Thus, assessments are not neutral but instead shape the way that teachers can teach (to varying degrees) and influence what students can learn.

As we reflect on this relationship (assessments, teaching, learning), we must consider *which* assessments produced pedagogical changes, *what* those pedagogical changes are doing, and *how* those pedagogical changes affect young people's learning

experiences. Regarding *which* tests produced changes, our results suggested fitness testing (not sport skills tests) spurred the changes in pedagogy. This was because students reported a change in *what* activities they engaged in: fitness training (e.g., running, muscular strength). Importantly, *how* students responded to these changes were not positive. Like previous research, students reported physical education as non-motivating (Jaakkola et al. 2016), unenjoyable (Alfrey and Gard 2014), painful (Wrench and Garrett 2008), and lacking cultural relevance (Safron and Landi 2022). Most importantly, fitness testing went against these young people's cultural beliefs about health, the body, and learning as holistic and multi-dimensional.

It is also important to mention the reactions these students described are unique (perhaps magnified) compared to previous research (outside China) because the results had serious consequences. As noted earlier, fitness testing literature has been equivocal, with studies exhibiting improved attitudes under specific conditions (Jaakkola et al. 2016; Simonton, Mercier and Garn 2019), positive attitudes amongst boys (Mercier and Silverman 2014; O'Keeffe, MacDonncha, and Donnelly 2021), a-motivation and low intrinsic motivation (Jaakkola et al. 2016; Goudas, Biddle and Fox 1994), as well as numerous studies documenting negative experiences (e.g., Hopple and Graham 1995; Wrench and Garrett 2008; Safron and Landi 2022). In our study, the students (males and females) overwhelmingly felt fitness testing produced anxiety, led to injuries, and prompted negative responses. Therefore, there is evidence to caution the use of performance-based fitness tests to evaluate students in high-stakes environments.

Alfrey's (2023) recent argument around the need to re-consider teaching practices around fitness to include student voice and articulate better educational aims

comes to matter here. Based on our findings, it seems the field of physical education would benefit from a re-articulation of testing and pedagogy that emphasises a holistic and multi-dimensional view of learning about health and the body (e.g., Quennerstedt, Landi and Casey 2024) including a focus on the affective domain (Kirk 2020). Further, a re-articulation must be considered in the construction of physical education assessments more broadly to consider whose bodies are excluded when specific physical performances are valued (e.g., team sport skills, power). A re-articulation of assessment from this point of view would mean moving away from performance or fitness mastery. Instead, assessments could be pedagogical tools that encourages students to move and express themselves in diverse ways, and perhaps even develop new and divergent forms of movement (Quennerstedt 2019).

Conclusion

Our goal was to explore Chinese students' reflective experiences of high-stakes assessment in physical education. We wanted to understand the perceived usefulness of these assessments for learning. What we found is that the high-stakes assessments focused solely on physical performance. The students reported that this singular emphasis stripped physical education of the holistic and multi-dimensional qualities that made it unique in schools. Thus, the tests often defeated what these students believed was the purpose of physical education. In so doing, they reported negative experiences, lack of student learning, and developing poor dispositions.

What students particularly disliked is when teachers shifted their pedagogical approach directly preceding high-stakes assessments to improve scores. Despite the test

being comprised of different outcomes (e.g., team sports, individual sports, fitness), all students reported their teachers focused solely on physical fitness training. As such, we found that high-stakes fitness tests can influence teachers' pedagogical decisions that centre movement around 'training' rather than learning. Notably, this had detrimental consequences for how students experienced physical education and did not align to what they valued about human movement.

Finally, students reported the test lacked diversity and failed to account for different body types and forms of health. As such, the test items often privileged physically fit and athletic bodies from a narrow perspective. The students who participated in non-traditional forms of movement or did not have bodies that 'fit' into an expected norm, felt excluded and bad about themselves. Thus, we advocate for assessments to be constructed in a way that considers the holistic nature of the body.

References

Alfrey, Laura. 2023. 'An Expansive Learning Approach to Transforming Traditional Fitness Testing in Health and Physical Education: Student Voice, Feelings and Hopes'. *Curriculum Studies in Health and Physical Education* 0 (0): 1–16. https://doi.org/10.1080/25742981.2023.2183477.

Alfrey, Laura, and Michael Gard. 2014. 'A Crack Where the Light Gets in: A Study of Health and Physical Education Teachers' Perspectives on Fitness Testing as a Context for Learning about Health'. *Asia-Pacific Journal of Health, Sport and Physical Education* 5 (1): 3–18. https://doi.org/10.1080/18377122.2014.867790.

Alfrey, Laura, and Dillon Landi. 2023. 'Fitness Testing as a Debated and Contested PE-for-Health Practice'. In *Physical Education Pedagogies for Health*, edited by Lorraine Cale and Jo Harris, 33–47. London: Routledge.

Alvesson, Mats, and Kaj Sköldberg. 2000. *Reflexive Methodology*. London: SAGE Publications.

611 Cale, Lorraine, and Jo Harris. 2009. 'Fitness Testing in Physical Education: A
612 Misdirected Effort in Promoting Healthy Lifestyles and Physical Activity?'
613 *Physical Education & Sport Pedagogy* 14 (1): 89–108.
614 https://doi.org/10.1080/17408980701345782.
615 Chen, Junjun, and Gavin TL Brown. 2013. 'High-Stakes Examination Preparation That
616 Controls Teaching: Chinese Prospective Teachers' Conceptions of Excellent
617 Teaching and Assessment'. *Journal of Education for Teaching* 39 (5): 541–56.
618 Chen, Xiaorong, Jiabao Cui, Ru Li, Richard Norton, Joel Park, Jian Kong, and Albert
619 Yeung. 2019. 'Dao Yin (a.k.a. Qigong): Origin, Development, Potential
620 Mechanisms, and Clinical Applications'. *Evidence-Based Complementary and*
621 *Alternative Medicine* 2019 (1): 3705120. https://doi.org/10.1155/2019/3705120.
622 Communist Party of China and State Council. 2007. 'Opinions on Strengthening Youth
623 Physical Education and Enhance Physical Fitness of Young People'. *Beijing:*
624 *CPC and State Council*.
625 Creswell, J. W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods*
626 *Approaches*. Thousand Oaks, CA: Sage.
627 Farvis, John, and Stephen Hay. 2020. 'Undermining Teaching: How Education
628 Consultants View the Impact of High-Stakes Test Preparation on Teaching'.
629 *Policy Futures in Education* 18 (8): 1058–74.
630 Göloglu Demir, Cennet, and Özlem Kaplan Keles. 2021. 'The Impact of High-Stakes
631 Testing on the Teaching and Learning Processes of Mathematics'. *Journal of*
632 *Pedagogical Research* 5 (2): 119–37.
633 Goudas, Marios, Stuart Biddle, and Kenneth Fox. 1994. 'Achievement Goal
634 Orientations and Intrinsic Motivation in Physical Fitness Testing With
635 Children'. *Pediatric Exercise Science* 6: 159–67.
636 Gu, Mingyuan, Jiansheng Ma, and Jun Teng. 2017. 'The Gaokao Experience of Chinese
637 Students'. In *Portraits of Chinese Schools*, edited by Mingyuan Gu, Jiansheng
638 Ma, and Jun Teng, 59–75. Singapore: Springer. https://doi.org/10.1007/978-981-
639 10-4011-5_4.

Hopple, Christine, and George Graham. 1995. 'What Children Think, Feel, and Know about Physical Fitness Testing'. *Journal of Teaching in Physical Education* 14 (4): 408–17. https://doi.org/10.1123/jtpe.14.4.408.

Hu, Tracy. 2017. 'Chinese Parents Feed Stimulants to Teens to Pass Entrance Exams'. *South China Morning Post*, 2017, sec. China Society. https://www.scmp.com/news/china/society/article/2102517/chinese-parents-feed-stimulants-teens-pass-senior-high-school.

Jaakkola, Timo Tapio, Arja Sääkslahti, Sami Yli-Piipari, Mika Manninen, Anthony Watt, and Jarmo Liukkonen. 2016. 'Student Motivation Associated with Fitness Testing in the Physical Education Context'. *Journal of Teaching in Physical Education* 32 (3): 270–86. https://doi.org/10.1123/jtpe.32.3.270.

Jing, Xiaoyan, and Li Liu. 2019. 'The Equity of Gaokao (National University/College Entrance Examination) in China'. *Review of Educational Theory* 2 (3): 29–34. https://doi.org/10.30564/ret.v2i3.881.

Jin, Aijing. 2013. 'Physical education curriculum reform in China: A perspective from physical education teachers'. *Physical Education and Sport Pedagogy* 18 (1): 15-27.

Jiuzhang, Men, and Guo Lei. 2009. *A General Introduction to Traditional Chinese Medicine*. CRC Press.

João Pires, Manuel Duarte. 2019. 'Gaokao: Far More than an Exam'. *Diadorim* 21: 168–85. https://doi.org/10.35520/diadorim.2019.v21nEspa27418.

Keating, X.D. 2003. 'The Current Often Implemented Fitness Tests in Physical Education Programs: Problems and Future Directions'. *Quest* 55: 141–60.

Keating, X.D., Xiaolu Liu, Rachyl Stephenson, Jianmin Guan, and Michael Hodges. 2020. 'Student Health-Related Fitness Testing in School-Based Physical Education: Strategies for Student Self-Testing Using Technology'. *European Physical Education Review* 26 (2): 552–70. https://doi.org/10.1177/1356336X19874818.

Keating, X.D., and Stephen Silverman. 2009. 'Determinants of Teacher Implementation of Youth Fitness Tests in School-Based Physical Education Programs'. *Physical*

Education & Sport Pedagogy 14 (2): 209–25. https://doi.org/10.1080/17408980801974945.

Kirk, David. 2020. Precarity, Critical Pedagogy, and Physical Education. Abingdon: Routledge.

Koro-Ljungberg, Mirka, Diane Yendol-Hoppey, Jason Jude Smith, and Sharon B. Hayes. 2009. '(E)Pistemological Awareness, Instantiation of Methods, and Uninformed Methodological Ambiguity in Qualitative Research Projects'. *Educational Researcher* 38 (9): 687–99. https://doi.org/10.3102/0013189X09351980.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago press.

Kvale, Steinar. 1996. *InterViews: An Introduction to Qualitative Research*. Thousand Oaks, CA: SAGE Publications.

Landi, Dillon. 2023. 'Precarity, Fitness Testing and Critical Pedagogy: A Response-Able Approach'. Motrivivência 35 (6): 1–15. https://doi.org/10.5007/2175-8042.2023.e95538.

Landi, Dillon. 2024. 'Thinking qualitatively: Paradigms and design in qualitative research'. Qualitative research and evaluation in physical education and sport pedagogy. 31-47.

Landi, Dillon, Jennifer Walton-Fisette, and Sue Sutherland. 2021. 'Physical Education Policy Research in the United States: Setting a New Orientation'. Quest. https://doi.org/10.1080/00336297.2020.1866042.

Leavy, Patricia. 2017. Research Design. New York, NY: The Guilford Press.

Li, Xin. 2023. 'Zhongkao, Not Gaokao, Now the Make-or-Break Exam, Parents Say'. *Sixth Tone*, 23 June 2023, sec. News: Education. https://www.sixthtone.com/news/1013176.

López-Pastor, Víctor Manuel, David Kirk, Eloisa Lorente-Catalán, Ann MacPhail, and Doune Macdonald. 2013. 'Alternative Assessment in Physical Education: A Review of International Literature'. Sport, Education and Society 18 (1): 57–76.

Marshall, C, and G B Rossman. 2006. *Designing Qualitative Research*. Thousand Oaks, CA: Sage Publications.

McDonald, Samantha M., and Stewart G. Trost. 2015. 'The Effects of a Goal Setting Intervention on Aerobic Fitness in Middle School Students'. *Journal of Teaching in Physical Education* 34 (4): 576–87. https://doi.org/10.1123/jtpe.2014-0138.

Meng, Xuan, Andrew Horrell, Paul McMillan, and Guorong Chai. 2021. '"Health First" and Curriculum Reform in China: The Experiences of Physical Education Teachers in One City'. *European Physical Education Review* 27 (3): 595–612. https://doi.org/10.1177/1356336X20977886.

Mercier, Kevin, and Stephen Silverman. 2014. 'High School Students' Attitudes Toward Fitness Testing'. *Journal of Teaching in Physical Education* 33 (2): 269–81. https://doi.org/10.1123/jtpe.2013-0153.

Morgan, David L. 2014. 'Pragmatism as a paradigm for social research'. *Qualitative inquiry* 20 (8): 1045-1053.

Ni, Dandan, and Shiyu Zhang. 2021. 'The Latest Front in China's Battle for School Places: Gym Class'. *Sixth Tone*, 2021, sec. Features. https://www.sixthtone.com/news/1006802.

O'Keeffe, Brendan T., Ciaran MacDonncha, and Alan E. Donnelly. 2021. 'Students' Attitudes towards and Experiences of the Youth-Fit Health-Related Fitness Test Battery'. *European Physical Education Review* 27 (1): 41–56. https://doi.org/10.1177/1356336X20917416.

Pluim, Carolyn, and Michael Gard. 2018. 'Physical Education's Grand Convergence: Fitnessgram ®, Big-Data and the Digital Commerce of Children's Health'. *Critical Studies in Education* 59 (3): 261–78. https://doi.org/10.1080/17508487.2016.1194303.

Quennerstedt, Mikael. 2019. 'Physical Education and the Art of Teaching: Transformative Learning and Teaching in Physical Education and Sports Pedagogy'. *Sport, Education and Society* 24 (6): 611–23. https://doi.org/10.1080/13573322.2019.1574731.

Quennerstedt, Mikael, Dillon Landi, and Ashley Casey. 2024. 'Busier, Happier, and Good(Er) – 40 Years on from "Busy, Happy, and Good" as Success in Teaching

Physical Education'. *Quest*, 1–21. https://doi.org/10.1080/00336297.2024.2393624.

Randall, William L., and Cassandra Phoenix. 2009. 'The problem with truth in qualitative interviews: reflections from a narrative perspective'. *Qualitative research in sport and exercise* 1 (2): 125-140.

Ringrose, Jessica, and Rebecca Coleman. 2013. 'Looking and Desiring Machines: A Feminist Deleuzian Mapping of Bodies and Affects'. In *Deleuze and Research Methodologies*, edited by Rebecca Coleman and Jessica Ringrose, 125–44. Edinburgh: Edinburgh University Press.

Rubin, Herbert J., and Irene Rubin. 2005. *Qualitative Interviewing: The Art of Hearing Data*. London: SAGE Publications.

Ryan, Katherine. 2002. 'Assessment Validation in the Context of High-stakes Assessment'. *Educational Measurement: Issues and Practice* 21 (1): 7–15.

Safron, Carrie, and Dillon Landi. 2022. 'Beyond the BEEPs: Affect, FitnessGram®, and Diverse Youth'. *Sport, Education and Society* 27 (9): 1020–34. https://doi.org/10.1080/13573322.2021.1953978.

Saldaña, Johnny. 2013. *The Coding Manual for Qualitative Researchers*. 2nd ed. Thousand Oaks, CA: SAGE Publications.

Shanghai Education Municipal Committee. 2019. 'Notice of the Shanghai Municipal Education Commission on the Issuance of the Implementation Plan of the Shanghai Junior High School Graduation and Promotion physical education Examination' *Shanghai Education Commission Regulation.* https://edu.sh.gov.cn/zcjd_area_3688/20200706/0015-xw_100194.html

Shuo, Zuo. 2020. 'Provinces Increase Emphasis on PE Scores'. *China Daily*, 2020, sec. Education. https://www.chinadaily.com.cn/a/202012/15/WS5fd80919a31024ad0ba9bced.html.

Smith, Brett, and Kerry R. McGannon. 2018. 'Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology'. *International review of sport and exercise psychology* 11(1): 101-121.

761 Smith, Brett, and Cassandra Phoenix. 2019. 'Qualitative Research in Physical Activity
762 and Health'. In *Research Methods in Physical Activity and Health*, edited by
763 Stephen R. Bird, 109–16. Abingdon, Oxon: Routledge.
764 Smith, Brett, Andrew C. Sparkes, and Nick Caddick. 2014. 'Judging Qualitative
765 Research'. In *Research Methods in Sports Coaching*, edited by Lee Nelson,
766 Ryan Groom, and Paul Potrac, 192–201. Abingdon, Oxon: Routledge.
767 Silverman, Stephen. 2017. 'Attitude Research in Physical Education: A Review'.
768 *Journal of Teaching in Physical Education* 36 (3): 303–12.
769 https://doi.org/10.1123/jtpe.2017-0085.
770 Silverman, Stephen, X.D. Keating, and Sharon R. Phillips. 2008. 'A Lasting
771 Impression: A Pedagogical Perspective on Youth Fitness Testing'. *Measurement*
772 *in Physical Education and Exercise Science* 12 (3): 146–66.
773 https://doi.org/10.1080/10913670802216122.
774 Simonton, K., Kevin Mercier, and Alex C. Garn. 2019. 'Do Fitness Test Performances
775 Predict Students' Attitudes and Emotions toward Physical Education?' *Physical*
776 *Education and Sport Pedagogy* 24 (6): 549–64.
777 https://doi.org/10.1080/17408989.2019.1628932.
778 Strom, Kathryn J., and Adrian D. Martin. 2017. *Becoming-Teacher: A Rhizomatic Look*
779 *at First-Year Teaching*. Rotterdam, The Netherlands: Sense Publishers.
780 Sun, Da-zhi, Shao-dan Li, Yi Liu, Yin Zhang, Rong Mei, and Ming-hui Yang. 2013.
781 'Differences in the Origin of Philosophy between Chinese Medicine and
782 Western Medicine: Exploration of the Holistic Advantages of Chinese
783 Medicine'. *Chinese Journal of Integrative Medicine* 19 (9): 706–11.
784 https://doi.org/10.1007/s11655-013-1435-5.
785 Supovitz, Jonathan. 2009. 'Can High Stakes Testing Leverage Educational
786 Improvement? Prospects from the Last Decade of Testing and Accountability
787 Reform'. *Journal of Educational Change* 10 (2): 211–27.
788 Tsai, Chiung-Tzu Lucetta, and Lijun Zhou. 2017. 'The history of sports and physical
789 education in China'. *World leisure journal* 59 (4): 326-337.
790 Wang, Lijuan, Amy Sau-ching Ha, and Xu Wen. 2014. 'Teaching Perspectives of
791 Chinese Teachers: Compatibility with the Goals of the Physical Education

Curriculum'. *Journal of Teaching in Physical Education* 33 (2): 213–31. https://doi.org/10.1123/jtpe.2013-0055.

Whittle, Rachael Jayne, Amanda Clare Benson, and Amanda Telford.2017 'Enrolment, content and assessment: a review of examinable senior secondary (16–19 year olds) physical education courses: an international perspective'. *The Curriculum Journal* 28 (4): 598-625.

Wrench, Alison, and Robyne Garrett. 2008. 'Pleasure and Pain: Experiences of Fitness Testing'. *European Physical Education Review* 14 (3): 325–46. https://doi.org/10.1177/1356336X08095669.

Wu, Yingkang. 2015. 'The Examination System in China: The Case of Zhongkao Mathematics'. In, 897–914. Springer.