ECNU Review of Education
2021, Vol. 4(2) 396–409

DOI: 10.1177/2096531120944522
journals.sagepub.com/home/roe

Education Quality Assessment in China: What We Learned From Official Reports Released in 2018 and 2019

Danqing Yin (殷丹青)
University of Kansas

Abstract

Purpose: This article intended to introduce the recent Chinese compulsory education quality assessment.

Design/Approach/Methods: The article summarized education policy documents relevant to the recent Chinese compulsory education quality assessment, especially from the three reports: *The Chinese Compulsory Education Quality Monitoring Report*, *The Chinese Compulsory Education Quality Monitoring Report on Mathematical Learning*, and *The Chinese Compulsory Education Quality Monitoring Report on Physical Education and Health*.

Findings: The article examined the development of Chinese education quality assessment in the context of the nation's pursuit of holistic education. It first reviewed a brief history of the education quality monitoring work in China that bore the fruits of a few recent reports. Second, findings and the framework of the assessment are demonstrated. Finally, I discuss the significance, limitations, and special issues of the assessment.

Originality/Value: The article is among the earliest to present findings of the recent quality assessment in Chinese compulsory education in English critically.

Corresponding author:
Danqing Yin, Department of Educational Leadership and Policy Studies, School of Education and Human Sciences, University of Kansas, 1122 W Campus Rd, Lawrence, KS 66045, USA.
Email: danqing.yin@ku.edu

Keywords

Compulsory education, education quality assessment, large-scale assessment

Date received: 27 February 2020; accepted: 2 July 2020

The Chinese Compulsory Education Quality Monitoring Report was first published in July 2018 by the National Assessment Center for Education Quality in China, under the supervision of the Ministry of Education, People's Republic of China (hereinafter referred to as "Chinese MOE" or "MOE"). Two additional subject reports, *The Chinese Compulsory Education Quality Monitoring Report on Mathematical Learning* and *The Chinese Compulsory Education Quality Monitoring Report on Physical Education and Health*, were later released in 2019. These reports presented the assessment results from the 2015–2017 assessment cycle in a recent large-scale assessment in Chinese compulsory education.

This article aims to examine the development of Chinese education quality assessment in the context of prevailing international large-scale assessments and the nation's pursuit of holistic education. It first reviewed a brief history of the education quality monitoring work in China that bore the fruits of the above three reports. The general information of Chinese compulsory education, the guiding law, and the policy for the assessment are introduced in this section. Second, findings from the reports are presented by individual report with a comparative analysis on results of available subjects: One is mathematics, and the other is physical education and health. The framework of the assessment is also demonstrated. Finally, I discuss the significance and limitations of the compulsory education quality monitoring in China. Special attention is placed on the common issues that international large-scale assessments have shared and the unique issues of Chinese assessment.

There has not been a consistent English translation for the Chinese names of reports and assessment work. Chinese MOE translated the 2018 report as the "China Compulsory Education Quality Oversight Report" and the 2019 subject reports as the "National Assessment of Education Quality—Mathematics" and the "National Assessment of Education Quality—PE and Health" (The National Assessment Center for Education Quality, 2018, 2019a, 2019b). Some researchers referred the assessment work as the "National Assessment of Education Quality" and the subject report as the "China National Assessment of Education Quality—Physical Education & Health"; some used "education monitoring" or "educational surveillance" to represent education quality assessment (Jiang et al., 2019; Tian & Sun, 2018; Wu et al., 2019).

To speak with a larger and broader audience who come from different countries, and to indicate the stage of the education that was originally included in the names of Chinese reports, the author avoided ambiguous adjectives like the "national" and applied following names, which reflected the assessment work and content of the reports more accurately and consistently:

1. *The Chinese Compulsory Education Quality Monitoring Report,*
2. *The Chinese Compulsory Education Quality Monitoring Report on Mathematical Learning, and*
3. *The Chinese Compulsory Education Quality Monitoring Report on Physical Education and Health.*

This policy review uses education assessment and education monitoring interchangeably to identify the related work since 2015 that generated the recent reports for clarification purposes. Subject and subdisciplinary are also used interchangeably in this review.

Background of the *Chinese Compulsory Education Quality Monitoring Work*

According to China's Ministry of Education, as of 2015, China has more than 260,000 compulsory schools (covering Grade 1 through Grade 9), more than 9 million full-time teachers, and nearly 140 million students (MOE, 2015a, 2015b). Policy documents showed that the major purpose of monitoring the quality of compulsory education is

> to scientifically assess the overall quality of compulsory education in China, objectively unfold the basic situation of relevant factors that affect the quality of compulsory education, and systematically monitor the implementation of national curriculum standards, eventually providing reference to educational improvements. (MOE, 2015a, 2015b)

The overall assessment work related to these reports is guided by *The Compulsory Education Law of the People's Republic of China* and *The National Compulsory Education Quality Monitoring Policy*, with a view to assessing the progress of the compulsory education and improving the quality of education with effective policies. Both of these legal instruments, first initiated in 1980s and 2010s, respectively, rooted in *The Constitution of the People's Republic of China* and intended to protect the rights to receive basic education for children of school age. As *The Compulsory Education Law of the People's Republic of China* in 1986 articulated, a 9-year compulsory education should be implemented by Chinese governments at different levels (6- to 15-year-olds, Grades 1–9), mainly serving primary school students and middle school (also called junior high school) students.

Before officially launching the recent compulsory education quality assessment, China had already possessed a few conducive conditions for it (Jiang et al., 2019; MOE, 2015a, 2015b; Wu et al., 2019).

Regarding the leading organization, the National Assessment Center for Education Quality (hereinafter referred to as "the Center") was formally established and housed in Beijing Normal University, a public and prestigious research university located in Beijing, China, with a 118-year

history so far (BNU History). Since the establishment in November 2007, the Center has been responsible for performing daily work of assessment.

Led by the Center, approximately 300 domestic experts, international scholars, special personnel of leading departments, principals of primary and secondary schools, and schoolteachers—a capable group of representatives—participated in the assessment work and supervised the assessment.

The experts developed multiple research tools for the assessment in different subject areas and disciplines. For example, paper test tools and field test tools were designed to measure the performance of six subjects including moral education, Chinese reading, mathematics, science, art, and physical education.

Furthermore, a pilot study for the quality assessment had been conducted in those six subject areas for eight consecutive years, five of which were nationwide large-scale tests with the sample from 32 Chinese provinces, municipalities, autonomous regions, and other equivalent units. In total, more than 460,000 students, 110,000 teachers, and principals from 695 sample counties (cities and districts) participated in the pilot study.

In a nutshell, a top-down administrative approach had been tested and moderated by trials and errors, with the Chinese MOE supervising the assessment, the Center organizing the assessment, the participated provinces coordinating the assessment, and counties further operating the assessment. Specific methods designed included sampling, testing, data analysis, and so forth. All of these work provided basic conditions and practical experience for the Ministry of Education to carry out a larger scale of compulsory education quality monitoring throughout China.

Findings from *The Chinese Compulsory Education Quality Monitoring Report*

The nationwide project that generated *The Chinese Compulsory Education Quality Monitoring Report* (hereinafter referred to as "the 2018 Report") officially started in 2015. Evaluations were conducted in mid-June each year from 2015 to 2017. During this first cycle of quality assessment, a total of 572,314 fourth-grade and eighth-grade students from 32 Chinese provinces, municipalities, autonomous regions, and other equivalent units have participated in the assessment, where a broad range of knowledge and skills, processes and methods, emotional attitudes and values were measured. In addition, 19,346 primary and middle school principals and 147,610 schoolteachers completed the questionnaire survey.

According to the 2018 Report, the large-scale education quality monitoring produced the following sample results.

For moral education, the questions were mostly around cultural, artistic, social, and historical achievements of the country. The result showed that a majority of Chinese students had a positive

life value orientation and had a good understanding of their traditional culture. To be specific, when asked whether they are proud of being a Chinese citizen, 96.2% of fourth graders and 97.9% of eighth graders said yes. The rate that sampled students answered cultural questions accurately and correctly was relatively high, with 66.1% among fourth graders and 74.5% among eighth graders.

As for Chinese reading, mathematics, and science subjects, for which learning usually happens inside a school building and for which higher stakes are frequently attached (Popham, 1987), sampled students did well. The typical monitoring questions categorized student academic performance into four levels, which include "to be improved," "medium," "good," and "excellent." As for reading performance, the proportion of students in the fourth and eighth grades reached and above "medium" were 81.8% and 79.6%, respectively; about 21.0% fourth graders and 22.7% eighth graders reached the "excellent" level. As for mathematics performance, the proportions of students in the fourth and eighth grades reached and above "medium" were 84.6% and 78.9%, respectively; about 23.8% fourth graders and 26.7% eighth graders reached the "excellent" level. As for science, 76.8% of fourth graders and 83.6% of eighth graders achieved the "medium" or higher level; those who reached the "excellent" level in science were 16.0% for fourth graders and 12.0% for eighth graders.

As for art and music, and physical education, where learning not always happens inside a classroom and for which lower stakes are attached as most are not included in Chinese National College Entrance Examination (*Gaokao*), the assessment results left much space for students to improve.

For health, the classification from Chinese *National Student Physical Health Standard* (revised in 2014) put student weight into four scales: normal weight, low weight, overweight, and obesity. The results from the 2018 Report showed that the proportions of normal weight were 74.2% for male students and 79.2% for female students in fourth grade; 78.2% and 80.6% for male and female students, respectively, in eighth grade. As for low weight scale, the numbers for male and female students were 8.4% and 7.0% in fourth grade, 5.1% and 2.1% in eighth grade, respectively. Regarding the overweight rate, 8.9% of male and 8.6% of female fourth graders and 8.2% of male and 11.1% of female eighth graders fell into this scale. In addition, the obesity scale included 8.5% of male fourth graders and 5.1% for female ones; and the numbers were 8.5% and 6.2% for male and female eighth graders, respectively. For some regions, the obesity problems were more prominent, with a rate over 15%. However, the specific information about those regions was not disclosed in this report.

Besides the student weight condition, the 2018 Report also summarized a number of other aspects of health issues, such as the cardiopulmonary function, vision, and sleep quality, among sampled students. For example, the data showed that the detected rates of poor vision in Grades 4

and 8 were 36.5% and 65.3%, respectively. For some regions, the numbers could reach as high as 60% for fourth graders and 80% for eighth graders. Detailed information was not given in this report.

For music and art, the monitoring assessed student performance in singing and listening, as well as knowledge and appreciation of art. Tested students are required to sing two songs, one required, the other elective chosen from a pool of 10 different songs, which all came from student music textbooks at school (Hu, 2018). The following figures (Figures 1 and 2) showed the percentage of fourth graders and eighth graders reached the "medium" benchmark and landed above it (including "medium," "good," and "excellent") in five different categories that the assessment tested.

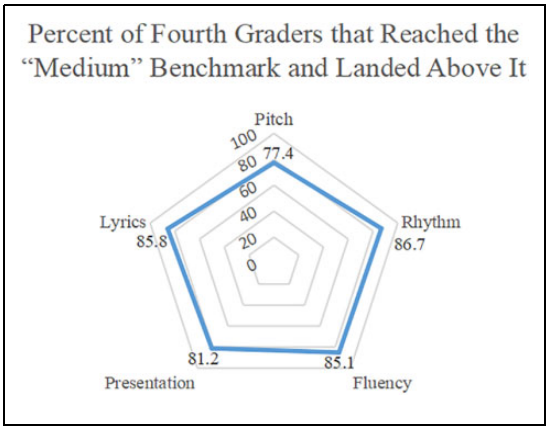

Figure 1. Percentage of the fourth graders who reached certain benchmark in music.
Source: The Chinese Compulsory Education Quality Monitoring Report, p. 14.

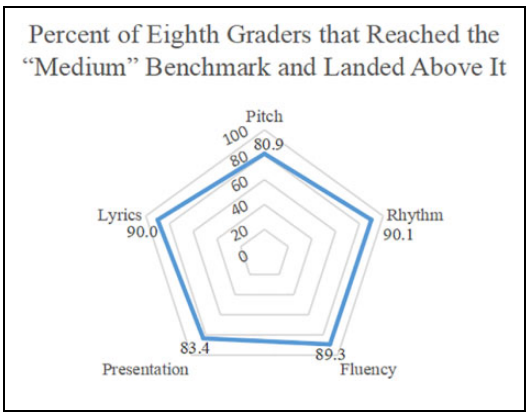

Figure 2. Percentage of the eighth graders who reached certain benchmark in music.
Source: The Chinese Compulsory Education Quality Monitoring Report, p. 14.

In addition, the assessment result showed correlations and suggestions. For instance, sampled students had a relatively high degree of interest in their curriculum, but the arrangements of course hours and course content could be more reasonable. Students generally spent too much time on homework, with a high percentage of off-campus tutoring classes and high learning pressure. Most teachers were liked by their students, but some teachers could be more responsible and professional. Plenty of schools have had positive cultural atmosphere and good environment, equipped with sufficient teaching facilities, which can be utilized more efficiently though. Parents generally cared about their children's learning and there was still room for improvement in parent–child communication and parenting type.

Results of the first subdisciplinary reports on the Chinese compulsory education quality assessment

The education quality monitoring in China has planned a 3-year cycle and a focus on two disciplines each year. By doing this, the assessment team intend to ensure the systematicity and consistency of the monitoring, to track changes of the quality of Chinese compulsory education, and to maximize the monitoring effectiveness. Specific arrangements have included the following: monitoring mathematics and physical education for the first year, language and art education for the second year, and science and moral education for the third year.

Based on *The Chinese Compulsory Education Quality Monitoring Report* released in July 2018, two monitoring disciplines, mathematics and physical education, were officially completed with subdisciplinary reports published, respectively, in the following year.

On November 20, 2019, the Center released two subdisciplinary results: *The Chinese Compulsory Education Quality Monitoring—Subject Report on Mathematical Learning* (2019a) and *The Chinese Compulsory Education Quality Monitoring—Subject Report on Physical Education* (2019b). These two subdisciplinary reports focused on presenting student learning achievement in mathematics and physical education with health condition and demonstrated takeaways from teachers working on related disciplines at school to further analyze factors that affect the quality of education.

The mathematics and the physical education subjects have been monitored in 32 Chinese provincial level units. A total of 6,680 primary and secondary schools got selected, with nearly 200,000 fourth graders and eighth graders being assessed (MOE, 2019). In addition, questionnaire surveys were conducted among more than 6,000 primary and secondary school principals and more than 30,000 mathematics and physical education teachers (MOE, 2019). The two reports used data to reveal the development of student performance in mathematics and physical education. As the fruit of the second-cycle monitoring, the two subdisciplinary reports released in 2019 not only summarized the specific monitoring results but also compared the results with those from the previous assessment cycle, displaying the yearly progress.

Following are details of the two subject reports that include the preparation of the related assessment, examples of specific assessment tools, general results from the assessment aligned by key questions, and some comparative results among different student groups.

Subdisciplinary report 1: The Chinese Compulsory Education Quality Monitoring on Mathematical Learning

Before the assessment of mathematics performance, more than 1,000 researchers and practitioners had engaged in the pilot discussions to develop the tools for the education quality assessment. For example, experts from the field of mathematics and pedagogy, education measurement and evaluation, education policy, and education leadership and management worked on finding valid tools for measurement; frontline teachers, principals, and researchers from Chinese primary and secondary schools participated in testing those tools and making improvement of the measurement (example of tools shown in Table 1). All the questionnaires put to use in real assessment process had already undergone multiple rounds of pretests, and each indicator was supposed to meet the requirement of metrology, with good reliability and validity.

As for the mathematics subject assessment, 116,529 fourth graders and 79,057 eighth graders participated. Among those responses, 115,804 effective ones from fourth graders and 77,949 ones from eighth graders were collected. The response rates were 99.4% and 98.6%, respectively. In addition, 12,505 fourth-grade math teachers, 12,438 eighth-grade math teachers, 4,139 primary school principals, and 2,538 junior high school principals participated effectively in the questionnaire survey, with a 100% response rate.

The results from this subject report in 2019 showed that students' mathematics performance made some progress, compared to the 2018 Report. For example, according to the subject report, the average mathematics score of the fourth graders was 502 points, which was 2 points higher than that from the previous report, while the latest monitoring result of the eighth graders showed an average mathematics score of 505 points, 5 points higher than the previous one.

While the overall performance of students in mathematics achieved a slightly better position, the proportions of students in Grades 4 and 8 that reached "medium" and higher levels in the measurement scale were 84.8% and 78.8% in 2018, respectively, which remained stable compared with their counterparts of 84.6% and 78.9%, respectively, in 2015.

As for motivation of learning mathematics, the proportion of students who indicated large interests in mathematics learning among fourth graders was 88.1%, and 67.9% for eighth graders in 2018. The fourth graders seemed having more confidence and less anxiety in mathematics: As 72.9% of them indicated having a "high" or "very high" level of self-confidence in mathematics learning, while the number was 58.8% among eighth graders in the same monitoring cycle on the one hand; on the other hand, the proportion of students indicating "low" or "very low" anxiety

Table 1. Assessment tools designed for mathematical learning assessment.

Assessment Categories	Assessment Indicators	Assessment Tools
Academic Performance	Arithmetic, Geometry, Data Analysis, Reasoning, Problem Solving	Mathematics Test (Paper and Pen)
Academic Attitudes and Emotions	Motivation of Mathematics Learning, Confidence of Mathematics Learning, Anxiety of Mathematics Learning	Student Survey (Paper and Pen)
Education and Teaching Condition	Popularity of Teachers, Education Degree of Teachers, Training Experience of Teachers, Internet Usage of Classes, Library Resources Equipment and Utilization, Multi-media Equipment, etc.	Student Survey (Paper and Pen), Teacher Survey (Online), and Principal Survey (Online)

Source: The Chinese Compulsory Education Quality Monitoring on Mathematical Learning, 2019a, p. 2.

level of learning mathematics was 75.2% among fourth graders and 59.1% among eighth graders of sampled students.

The 2019 subject report for mathematics also showed student satisfaction with their teachers, or teachers' popularity among their own students, as the percentage of fourth graders who adored their mathematics teachers was 92.2%, while that number of eighth graders was 85.6%. As for the percentage of mathematics teachers who met national basic degree qualification of teaching mathematics, the number was 97.1% for fourth grade and 99.4% for eighth grade, which remained steady since 2015 (with the numbers of 97.1% for fourth grade and 99.0% for eighth grade then).

According to the subject report in mathematics, the proportions of the fourth-grade mathematics teachers participating in various levels of professional development were between 41.0% and 93.7%, with more than 99% of mathematics teachers holding the belief that professional training is helpful, while the proportion of the eighth-grade mathematics teachers participating in professional development ranged from 40.4% to 91.4%, more than 97% of whom found the professional training helpful.

Subdisciplinary report 2: The Chinese Compulsory Education Quality Monitoring on Physical Education

To write *The Chinese National Compulsory Education Quality Monitoring—Subject Report on Physical Education*, the team assessed 116,631 fourth graders and 79,078 eighth graders from 4,141 primary schools and 2,539 junior high schools.

Furthermore, a total of 6,854 fourth-grade physical education teachers, 5,750 eighth-grade physical education teachers, 4,139 primary school principals, and 2,538 junior high school principals participated in the monitoring questionnaire surveys. The response rate was 100%. Results were presented in the subject report about physical education and health.

Results from the physical education subject report showed that the overall health condition of the students in the fourth and eighth grades was relatively good, as the rate of students reaching overall national physical fitness standards exceeded 93.0% and 85.0% for fourth graders and eighth graders, respectively. However, there were increasingly high obesity and poor-vision issues among students. For instance, the obesity rates of fourth graders were 8.8% and 9.7% for eighth graders, which were 1.9 and 2.2 percentage points higher than numbers from those student groups in 2015. The detected rates of poor vision in fourth graders and eighth graders were 38.5% and 68.8%, which were 2.0 and 3.5 percentage points higher than that in 2015.

In addition, as for sleep time, the 2019 report showed that 22.2% of students in the fourth grade could sleep 10 hr or more, the proportion of which decreased 8.5 percentage points from the 2018 report. The proportion of students in the eighth grade who spent 9 hr or more in sleep time was 19.4%, which was 2.8 percentage points higher than previous result. As for minimum requirement for class time of physical education, 69.2% of the fourth graders and 51.9% of the eighth graders had certain classes that met the national requirement of three sessions per week. These numbers had the rises of 13.5 percentage points and 12.7 percentage points, respectively, from that of previous result.

From the analysis of key factors on student physical performance, the subject report found that good physical exercise habits, healthy lifestyles, parents' appreciation and support for physical exercises, and proper implementation of physical education curriculum and teaching at school were all factors related to positive results on students' physical fitness.

The significance of the compulsory education quality assessment in China

China has a long history of holistic education that dated to the Confucius thinking and its philosophical product. Three thousand years ago, Chinese people in Zhou Dynasty highly valued "six arts" (*liuyi*): "rites, music, archery, charioteering, literature, and mathematics" for talent cultivation in Confucian aristocracy (Jiang et al., 2019). That was the early expression of a holistic education, the essence of which remained today for basic education in China.

With the hope to ensure students to receive quality education and grow holistically, the monitoring work added nontraditional subjects into assessment process, such as the art and music, physical education and health (Jiang et al., 2019; Tan, 2018). The embrace of those subjects in addition to reading and mathematics might not be an absolutely new attempt, regarding what an

education assessment can cover. Decades earlier than the recent assessment, there have been a number of subject areas covered for compulsory level education assessment in China of a smaller scale. For example, the Beijing Municipality Compulsory Education Quality Assessment Project in 1990s tested subject areas including Chinese, mathematics, arts, civics, handiwork, and common knowledge in natural science for primary schools (generally Grade 1 through Grade 6 or Grade 5); and for junior middle schools (Grade 7 or 6 through Grade 9), Chinese, mathematics, English, music, fine arts, humanities, natural sciences, hands-on skills, and behaviors have been covered in the education quality assessment (Tian & Sun, 2018, p. 187). However, it was adventurous and uncommon, considering the inclusiveness and the scopes of assessment are not seen in other countries, or international large-scale assessments, which only focus on few subjects.

Consequently, innovative methods were sought, applied, and invented. As examples, the stratified probability proportionate to size (PPS) sampling, systematic PPS sampling, and stratification were adopted in student sampling; partial balanced incomplete block design was employed for item development in assessment booklets; unidimensional item response theory models, the Rasch model, and the partial credit model were used for scaling and scoring assessment results. The experience from coordinating the complex quantitative and qualitative process of the assessment was valuable and transferable.

Moreover, assessment results offer empirical evidence for policy changes and social discussions. For example, based on specific and comprehensive profile of students living in high poverty regions, the results have been incentivizing policy suggestions on more financial assistance, such as improved meal plans, extra fundings, and boarding fee subsidies (Tan et al., 2018). Applying policy changes helps the country to allocate educational resources based on real needs and to ensure more educational equity. Another example from the mathematics subject assessment, the report presented a statistical analysis of student learning attitude and their academic performance, arguing that students with more positive mathematical learning attitudes were correlated with higher academic performance in mathematics, which stimulated questions among Chinese society, such as "how to improve mathematical learning attitudes in China," "who has more positive mathematics learning attitudes," "who has better mathematics performance," "are there causal relations between these," and so forth.

Additional significance of the assessment lies in its dynamic application of new technology for assessment per se and for results display. The compulsory education quality monitoring work has relied on a range of new media and online platform to inform the public about the whole process. It uses both websites and mobile applications to monitor process and results, helping both the participants and the public understand it. The monitoring center has also launched a special network platform called "the National Compulsory Education Quality Monitoring Platform" (https://eachina.changyan.cn/), which displays the information from the monitoring center, for provincial level units and sample schools. The website also provides a variety of guiding documents for downloads, such

as operating methods and contact information for additional information of the monitoring. Furthermore, the team has also established a Chinese WeChat account called "Quality Monitoring and Big Data" that has regularly published articles on assessment topics.

To sum up, including broad subjects, applying advanced method, informing critical policy discussions, and engaging the public with new media helped the latest assessment stand out from traditional ones and become a significant Chinese national topic in education.

The limitation and caveats of Chinese compulsory education quality assessment

As large-scale assessments have exerted power over education policy research around the world to use it more responsibly, one should also be aware of the caveats of using monitoring results to inform policy. There are a few remained concerns regarding the assessment, including the limitation of large-scale assessment per se, the danger of using results for accountability purposes, and the lack of detailed evidence on real effects on policy, and the need of data disclosure.

There are inherent limitations of what large-scale assessment like this could measure. According to the researchers who work at the assessment Center, the designs and instruments of the recent education quality assessment could be learned from the experience of other advanced large-scale assessment programs like international assessment PISA and quality assessment from other advanced economies, such as the National Assessment of Educational Progress (NAEP) in the United States.

A number of policy researchers shared key insights of what large-scale assessment cannot measure. For example, these assessments cannot measure unknown constructs, thus cannot fully predict success of individuals and societies; they cannot measure "exceptionality" and ignored the real frontiers and exceptional talents among top performers, thus stifling innovation; they cannot measure inconclusive, dynamic, and context-dependent human abilities, such as effective communication skills in different cultural contexts; assessments do not represent the qualities that different individual needs merely using the same standards to measure every child (Pons, 2017; Yuan & Zhao, 2019; Zhao et al., 2019).

As for reflecting the real purposes of education, previous critique argues that what large-scale assessments tried to represent is still too simplified that it ignored other meaningful educational purposes such as "citizenship, solidarity, equity, curiosity and engagement, compassion, empathy, curiosity, cultural values, physical health," and so on (Zhao, 2019). These are just a few examples of assessment concerns. Without fully addressing concerns like this, using the same standards to measure individuals' ability and make policy changes can be improper and prejudiced.

Furthermore, it can be dangerous to make hasty comparison and hold any provincial unit accountable. Researchers worldwide cast doubts on making simple comparison merely using rankings in assessment results to judge the education quality. Looking inside of a nation's

education systems could be proper and helpful, compared to looking outside from other countries due to different cultural and social backgrounds (Carnoy et al., 2015). However, policymakers should also be careful about how much variance actually happened within different economic, cultural, and social backgrounds of these units.

Next, unlike NAEP in the United States that published open data for the public and researchers to use and examine (see https://nces.ed.gov/nationsreportcard/data/), the data from the Chinese compulsory education assessment collected from provinces and counties have been kept exclusively and confidentially. There is no way for researchers outside of the Center and for researchers at Beijing Normal University to reach the data, conduct statistical analysis, or examine the national analysis published in national reports.

In addition to the inherent limitation of quality assessments and dangerous comparison using those results, there are lack of direct research on the positive impact of such quality assessment on making education policy, neither are any detailed data open to researchers to use and examine. So far, what we really know about the effect and significance of the Chinese compulsory education quality monitoring work are still fairly limited.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Beijing Normal University (BNU). (2020). *BNU history.* https://english.bnu.edu.cn/about/bnuhistory/index.htm

Carnoy, M., Garcia, E., & Khavenson, T. (2015). *Bringing it back home: Why state comparisons are more useful than international comparisons for improving US education policy.* Economic Policy Institute. https://www.epi.org/publication/bringing-it-back-home-why-state-comparisons-are-more-useful-than-international-comparisons-for-improving-u-s-education-policy/

Compulsory Education Law of the People's Republic of China. (2014). *Laws (and) regulations.* The State Council, People's Republic of China. http://english.www.gov.cn/archive/laws_regulations/2014/08/23/content_281474983042154.htm

Hu, Y. (2018). Analysis on the way of singing and music assessment in the first round of Chinese compulsory education quality monitoring [in Chinese]. *Teacher Doctrines*, (8), 256–257. http://m.fx361.com/news/2018/0719/3861833.html

Jiang, Y., Zhang, J., & Xin, T. (2019). Toward education quality improvement in China: A brief overview of the national assessment of education quality. *Journal of Educational and Behavioral Statistics*, *44*(6), 733–751.

Ministry of Education of the People's Republic of China. (2015a). *MOE Official Press Conference Material Part I. Establishing a national monitoring system to promote the quality of compulsory education* [in Chinese]. http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/xwfbh_2015n/xwfb_150415/150415_sfcl/201504/t20150415_187149.html

Ministry of Education of the People's Republic of China. (2015b). *MOE Official Press Conference Material Part II. Spokesperson from Chinese Ministry of Education answering reporters' questions on the National Compulsory Education Quality Monitoring Program* [in Chinese]. http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s8677/201504/185888.html

Ministry of Education of the People's Republic of China. (2018). *Release of China's first oversight report on quality of compulsory education*. http://en.moe.gov.cn/News/Top_News/201808/t20180801_344002.html

Ministry of Education of the People's Republic of China. (2019). *Release of China's national compulsory education quality monitoring subject reports on mathematics and PE* [in Chinese]. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201911/t20191120_409046.html

Pons, X. (2017). Fifteen years of research on PISA effects on education governance: A critical review. *European Journal of Education*, *52*(2), 131–144.

Popham, W. J. (1987). Preparing policymakers for standard setting on high-stakes tests. *Educational Evaluation and Policy Analysis*, *9*(1), 77–82.

Tan, H. (2018). On changes of quality monitoring of basic education in the new era [in Chinese]. *Educational Research*, *461*(6), 98–104.

Tan, H., Li, W., & Luo, L. (2018). Considerations on using quality assessment to promote targeted poverty alleviation in basic education [in Chinese]. *Educational Research*, *456*(1), 99–107.

The National Assessment Center for Education Quality. (2018). *The China compulsory education quality oversight report* [in Chinese]. http://www.moe.gov.cn/jyb_xwfb/moe_1946/fj_2018/201807/P020180724685827455405.pdf

The National Assessment Center for Education Quality. (2019a). *The Chinese national compulsory education quality monitoring—Subject report on mathematical learning* [in Chinese]. http://www.moe.gov.cn/s78/A11/moe_767/201911/W020191120537554536852.pdf

The National Assessment Center for Education Quality. (2019b). *The Chinese national compulsory education quality monitoring—Subject report on physical education* [in Chinese]. http://www.moe.gov.cn/s78/A11/moe_767/201911/W020191120537554570333.pdf

Tian, H., & Sun, Z. (2018). *Academic achievement assessment: Principles and methodology*. Springer.

Wu, L., Ma, X., Shi, Y., Tao, S., Yu, Y., Wang, S., Luo, L., Xin, T., & Li, Y. (2019). China national assessment of education quality—Physical education & health (CNAEQ-PEH) 2015: An introduction. *Research Quarterly for Exercise and Sport*, *90*(2), 105–112.

Yuan, Z., & Zhao, Y. (2019). Respect the power of large-scale assessments: What they cannot measure. *ECNU Review of Education*, *2*(3), 253–261.

Zhao, Y., Wehmeyer, M., Basham, J., & Hansen, D. (2019). Tackling the wicked problem of measuring what matters: Framing the questions. *ECNU Review of Education*, *2*(3), 262–278.

Zhao, Y. (2019). Yong Zhao: The PISA illusion. National Education Policy Center. https://nepc.colorado.edu/blog/pisa-illusion