

Educational Assessment of the Post-Pandemic Age: Chinese Experiences and Trends Based on Large-Scale Online Learning

Hong Su, China National Institute of Education Sciences

Abstract: Owing to the break-out of the COVID-19 pandemic, students have to take more online learning than offline, and large-scale education assessment programs have to be suspended or postponed. How could education assessment adapt to large-scale online learning? How could the effect and safety of online assessment be improved? What role should formative assessment play in student admissions? How could different assessment results be linked? Reflections on and trends of the Chinese experiences are presented in this article. Based on cross-cultural comparison research, measures to be recommended are as follows: reviewing previous theories, improving existing methods continuously, and developing assessment techniques innovatively according to new application scenarios.

Keywords: formative assessment, large-scale, online assessment, online learning, pandemic age, post-pandemic

Educational Assessment of the Post-Pandemic Age: Chinese Experiences and Trends Based on Large-Scale Online Learning

Since the break-out of the COVID-19 pandemic, many industries have been impacted and influenced to different extents, especially those labor-intensive industries. Unquestionably, the influence on the education industry was huge. Hundreds of millions of students could not study in their classrooms as usual all over the world, and a great number of large-scale education assessment¹ programs have to be suspended or postponed. For example, the Chinese government has decided to postpone the National College Entrance Examination (referred to as Gaokao hereinafter) that has always been the top priority for all high school students. The number of students who have registered for Gaokao, which can determine their college admission and future development, has reached 10.71 million in 2020. These students will be tested in hundreds of thousands of standardized examination rooms throughout the country at the same time from July 7 to 8. It will be a great challenge for the government to protect so many students from COVID-19 and ensure their safety. We have to admit that the long-lasting pandemic situation has an unparalleled influence on our education system. Professionals in different disciplines have conducted wide discussions about school teaching and education assessment, which are not only focusing on applying education information technologies, but also on reforming assessment theories, examination system, and measurement techniques. Some

vital and urgent topics of the discussions are presented in this article. These topics may reflect an important tendency in education assessment development in the world.

How Could Education Assessment Adapt to Large-Scale Online Learning?

A nine-year compulsory education policy is implemented in China. According to the Ministry of Education (MoE, 2020), there were 212,600 compulsory education schools with 154 million students in China in 2019 (MoE, 2020). Since the break-out of the COVID-19, a great number of students could only study online. As of the beginning of April in 2020, the number of four-year higher education institutions that started the new semester online reached 1,454 all over the country, with 713,000 online courses and 1.18 billion students learning online. In the first quarter of 2020, MOOC, an online course platform in China, added another 5,000 courses, and other online education platforms added 180,000 courses. All these facts could indicate that an online learning experiment with a large scale and a profound influence is taking its place in China, although it is unexpected. Unsurprisingly, the pandemic situation has promoted the rapid development of online learning and made it popular in a faster way globally. For many students, especially those in developing countries, this is a totally new learning experience.

At present, many educators are concerned that the students may not be able to adapt to such teaching mode and maintain the learning quality as usual. Some researchers have conducted a national survey on online learning and found out that 47.7% of the students believe that their academic results will be compromised, while more than 90% of all students and their parents are satisfied with online teaching. This is unacceptable to any student or school, let alone a

This work was supported by a grant from the Key Programs of the China National Institute of Education Sciences (No. GYB2019007). China National Institute of Education Sciences, No. 46 Beisanhuan Zhonglu Haidian District, Beijing 100088, P. R. China, suh@nies.net.cn; suhongabc@126.com

country that has made the improvement of education quality as its strategic target for education development. Although online learning has been accepted by more and more people in society, can it achieve the same effectiveness as face-to-face teaching? This is a hot topic for the interested parties of online learning—students, parents, teachers, school principals, online learning platform operators, and policymakers. Many teachers and education experts have researched the ways to maintain the quality of online learning. One of the important aspects is to deeply localize the online learning practice by basing it on the actual conditions of China according to the requirements specified in the Quality Assurance of Online Learning Toolkit issued by APEC (2019). Currently, some outstanding progress has been made on this aspect. Systematic research and practical measures have been made to improve Internet speed, provide high-quality learning resources, and offer more personalized teaching plans. In such a new learning scenario, education assessment, as an extremely important link of the whole teaching process, has also been widely discussed in terms of its concepts, methods, and strategies. For example, how can the students' performances be rapidly and accurately assessed? And how can the more personalized teaching resources be promoted based on the assessment? To answer these questions, the methods and techniques regarding web-based item bank, adaptive tests, automatic scoring, and online result reporting shall be considered.

But the mature experience in western countries could not be directly used in China, just like what a wise man in ancient China said, "Owing to different natural environment, the same species of citrus growth in both sides of River Huai, rooting in the south for honey orange, but in the north for trifoliate bitter." Many modern education assessment theories and techniques are developed based on the education practices in western countries. Education in China, however, has its own characteristics. An example is the difference between the Gaokao and the SAT. In China, subjective questions are always used to assess students. In the Gaokao test forms, for example, there are many subjective multistep questions worth a large number of score points ranging from 5 to more than 10. The Chinese essay even has 60 marks.² By contrast, SAT questions are mostly standard multiple choice questions. Such difference has made it difficult to apply some advanced education assessment theories and techniques, such as Item Response Theory and the Cognitive Diagnostic Model in China, in a rapid way. When theories or techniques do not apply in practices, most likely the theories or techniques should be changed.

Some may ask: why do not we only use multiple choice questions to test students? Based on the accuracy and influence of assessment results, educators in China generally believe that subjective questions are more effective, which has also been recognized by the A-level reformers in the UK. Multistep questions worth a large number of marks are valuable because they reward the method in which the problem is solved by the student. It was suggested that a greater use of multistep questions in testing might help to better encourage critical thinking skills (Higton, 2012). In general, there is a long way to go before giving a full play to the education assessment techniques in the online learning scenario and helping the improvement of learning quality.

How Could the Effect and Safety of Online Assessment be Improved?

Education assessment methods must be adapted to the changed learning style, so as to promote the learning in a better way. At present, a lot of emerging online learning platforms have developed various assessment tools, but many of them have only changed the way of presenting the assessment by moving the questions from papers to screens and changing the assessors from teachers to machines. This is undoubtedly the achievement of technological advance, but it is not the intended purpose of online assessment and will not achieve outstanding effects in this way. Online assessment can take advantage of the modern network technology to transfer and analyze massive data quickly, resulting in the convenience and effectiveness of assessment on generating test papers and distribution, scoring, and reporting. Thereby it saves time, human labor, expenses, and other costs in various aspects when compared to traditional assessment. However, the ultimate goal should be realizing more individualized and effective learning.

Clearly, it is far from enough to simply digitize textbooks, tests, and assessment tools. All the resources relating to learning should be connected with a key matter, such as the intelligent item bank, according to the existing practices in China. However, there are still two major challenges about this trend: one is that where the huge amount and high-quality items come from. It should be noted that China is a large country with 282 million students participating in various examinations and the quantity of test items used every day is unimaginable. The other one is how we obtain the parameter values of these items. Without parameter values, test items are only piled together, although it will be at huge cost to obtain the parameter values of these items. Safety is a more urgent issue in online assessment. The safety we talk about here has two aspects. One is the security and reliability in transmitting assessment tools, testing response data and other results, which shall not be leaked. The other is to ensure that the students answer the questions honestly and to prevent and detect any cheating behaviors. For the former, some research institutions and high-tech companies in both China and other countries have resolved the key issues with modern network technologies. For the latter, however, no reliable and feasible technique is available at present.

In fact, if the online assessment is only used for learning diagnosis and improvement, it is welcomed by many schools since they can use it for homework, stage quizzes, and even final examinations. Although the assessment results do not have high stakes, efforts shall also be made to prevent students from plagiarizing answers or looking for answers on the Internet (Lv, 2019). For the examinations with large scales and high stakes, it is impossible to promote online assessment widely for many reasons. First of all, there are too many students participating in such examinations who are distributed in different places, and some places could not meet the hardware conditions of online assessment in a short term. Second, many people do not think that online assessment is scientific and fair enough. Finally, some questions with high marks could not be answered online since the answering process must be described clearly as required by test developers. These are the main reasons why some large-scale examinations in China have been postponed.

Recognition of the Functions of Formative Assessment

For a long time, China has been relying on the results of summative assessment to select students for higher grade schools or universities. After completing compulsory education, students must take part in large-scale standardized admission examinations to go to good high schools or universities. The examination results, or the marks, are almost the only factor determining whether they can go to good schools or not. There is a saying in China called “An exam determines the fate.” Though it is not fully correct, it illustrates the importance of the fateful examination. In China, the result of one single examination may determine if a student could be admitted by the university, which is quite different from many higher education institutions in the world. For example, in the United States, according to the National Association for College Admission Counseling, more than 10 factors could influence the admission processes of colleges and universities, including Grades in All Courses, Strength of Curriculum, Admission Test Scores (SAT and ACT), Essay or Writing Sample, Recommendation, Class Rank, and Extracurricular Activities, and there is no definite plan or specific combination of factors that will guarantee a student admission. Academic performance in high school has been the most important consideration in freshman admission decisions for decades (NACAC, 2019). Obviously, the Admission Test Scores (ACT or SAT) are not so important as the Gaokao result in China. What is more, more than 1,200 Accredited, Four-Year Colleges and Universities with ACT/SAT-Optional Policies for Fall 2021 Admission (Fair Test, 2020) and most recently, University of California approved changes to their standardized testing requirement for undergraduates (UC Office of the President, 2020).

However, due to the pandemic situation, the large-scale standardized examinations could not be conducted according to schedules, and no other alternative method is available in the short term. A lot of discussions have been made on how to break the deadlock. Some education experts start to criticize the way of using examination results. In their opinion, although the selection method of over-relying on the results of summative examinations could meet the demand of the society for fairness, the role of the summative assessment has to be reduced in student admission if it is up against serious challenges or unsustainability. Of course, it is impossible to fully abandon the standardized examination for the college admission in China at present. Rediscovering the value of formative assessment and improving its role in decisions with high stake have become an important reform trend. Comprehensive assessment is one of the cores of the new round of reform of the examination and enrollment system in China now.

The so-called Comprehensive is to hope that the higher education institutions could pay more attention to other factors than the results of the standardized examination, including the formative assessment results of students when they are studying in high schools. Education experts and policymakers in China have paid close attention to the ways of top universities in the world on selecting students, and have started to realize that the formative assessment could not only avoid the large-scale aggregation issue but could also assess a student in a more comprehensive way and predicate the academic performance of the student in the future more effectively (Zhang, 2016). However, there are a lot of opposed voices. On the one hand, formative assessment is conducted by different schools and districts, making it difficult to compare objectively and challenging the fairness of the assessment. On

the other hand, once the importance of formative assessment is increased in student admission, it will have high stakes, which may lead to grade inflation of the assessment results or fake application information of the students, damaging the reliability and validity of the assessment tools. These conditions have occurred in the past reforms. Maybe this is a world-wide problem. We can predict that the debate in this aspect will keep going on.

Trying to Link Different Assessment Results

China is a big country with many kinds of assessments. It especially has a long history of examinations. Some crucial examinations, such as the Gaokao, have an important role to play in changing the fates of individuals, maintaining social fairness, and keeping social stability. Such role has strengthened the whole society’s understanding of examinations and too many people deem examinations as a part of the social culture (often called “exam culture” in Chinese).

Many Chinese parents are enthusiastic about examinations, and this is one of the reasons why some examinations are unbelievably hot in China. Last year, two English assessment programs, the Key English Test and the Preliminary English Test from the UK, saw the breakdown of their registration systems because too many students were registering online at the same time. Now, China has hundreds of millions of students, who are potential consumers of various assessments. The huge market has led to various assessment programs and some of them have similar functions or even could be replaced by each other. In fact, several years ago, the policymakers of government agencies have realized the necessity to abolish some similar assessment programs.

In 2014, the China State Council initiated an important reform, and one of the key items of the reform was to develop a national foreign language assessment system. Then, the experts in China started to do this work by integrating the English assessments at different stages. To date, some surprising progress has been made. At the end of 2018, IELTS and Aptis became the first English tests linked to China’s Standards of English Language Ability. Completion of the linking project marks that China’s standards of English language ability is officially mapped to the international examination system (British Council, 2018). What is more, in December 2019, an important research project to map TOEFL iBT test scores onto the CSE levels was also completed successfully (ETS, 2019). Building a common ability scale to reduce assessments with similar functions has become especially important after the occurrence of the pandemic. Furthermore, if different assessment results could replace each other in use, they must be able to deal with the influences generated by some regional and local emergencies, and be welcomed by students because this way could reduce the cost of taking part in different examinations and provide students with more choices. Just like ACT and SAT, even though they have different constructs, their developers have worked together to complete a concordance study and it could provide a tool for finding comparable scores for college admission. In terms of methods, various different statistical and measurement models could be used to link different assessment tools or at least their results, but some people still have doubts on whether the assessments could replace each other in this way. For example, a province in east China is implementing a reform in its high schools to provide all students with the

opportunity to take the examination twice a year for their English subject. Although every student could use the higher score of the two examinations to apply for universities, it is found that almost all students have participated in two examinations instead of one. In their eyes, this is an opportunity to get higher scores. As we all know, it is extremely difficult to initiate any reform on Gaokao in China, but the calls for linking low-risk assessments are becoming more clear.

By June 2020, China has made an excellent achievement in controlling the pandemic situation. Many schools have resumed their operations. Under various protection measures, students begin to return to classrooms. However, many after-school courses shall still be provided online. The combination of online and offline learning would be the product under the profound influence of the pandemic on the education system, which may become a new teaching and learning form in future pandemics. This is a common experience and a critical moment for initiating reforms. The education assessment industry needs this opportunity badly to eliminate the false and retain the true and seek common ground while reserving differences, so as to further the development of education assessment discipline and expand the market for assessment products. Some proposed measures that can be taken include: reviewing previous theories, improving existing methods continuously, and developing assessment techniques innovatively according to new application scenarios. Of course, all of these measures must be taken based on cross-cultural comparison research.

Notes

¹In Chinese, there are several words to describe the methods to evaluate students. In this article, assessment is used for convenience and it may include national large-scale tests and other examinations conducted by schools, universities, and districts.

²The full marks of the paper is 150 and the result will be reported as a raw score.

References

- British Council. (2018). Results of linking IELTS and Aptis to China's standards of english language ability. Retrieved at: <https://www.britishcouncil.cn/en/exams/cse/results>.
- Clinedinst, M. (2019). 2019 State of College Admission. Retrieved at: https://www.nacacnet.org/globalassets/documents/publications/research/2018_soca/soca2019_all.pdf
- ETS. (2019). China TOEFL® 2019 Annual Conference Held in Beijing. Retrieved at: <https://news.ets.org/press-releases/china-toefl-2019-annual-conference-held-in-beijing/>.
- Fair Test. (2020). Four-years colleges and universities will be test-optional for Fall 2021 admission. Retrieved at: <https://www.fairtest.org/more-half-all-us-fouryears-colleges-and-universiti>.
- Higton, J, Noble, J, Pope, S, Boal, N, Ginnis, S, Donaldson, R, & Greevy, H. (2012). Fit for purpose? The view of the higher education sector, teachers and employers on the suitability of A levels. Retrieved at: <http://dera.ioe.ac.uk/14104/1/2012-04-03-fit-for-purpose-a-levels.pdf>.
- Lv, X. (2019). The application of the anti-cheating measures in the online examination system. *Technology Wind*, 36(12), 65.
- Ministry of Education of the People's Republic of China (MoE). (2020). National Statistical Bulletin of Education Development in 2019. Retrieved at: http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202005/t20200520_456751.html.
- UC Office of the President. (2020). University of California Board of Regents unanimously approved changes to standardized testing requirement for undergraduates. Retrieved at: <https://www.universityofcalifornia.edu/press-room/university-california-board-regents-approves-changes-standardized-testing-requirement>.
- University of Melbourne (2019). APEC Quality Assurance of Online Learning Toolkit. Retrieved at: <https://www.apec.org/Publications/2019/12/APEC-Quality-Assurance-of-Online-Learning-Toolkit>.
- Zhang, J. (2016) A comparative study of statistical models for moderating school-based exam scores in college admissions. *Tsinghua Journal of Education*, 37(1), 93–98.

Copyright of Educational Measurement: Issues & Practice is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.