

2020

The Development of Psychological and Educational Measurement in China

Houcan Zhang

Fang Luo

Follow this and additional works at: <https://www.ce-jeme.org/journal>

Recommended Citation

Zhang, Houcan and Luo, Fang (2020) "The Development of Psychological and Educational Measurement in China," *Chinese/English Journal of Educational Measurement and Evaluation | 教育测量与评估双语季刊*: Vol. 1 : Iss. 1 , Article 7.

Available at: <https://www.ce-jeme.org/journal/vol1/iss1/7>

This Article is brought to you for free and open access by Chinese/English Journal of Educational Measurement and Evaluation | 教育测量与评估双语季刊. It has been accepted for inclusion in Chinese/English Journal of Educational Measurement and Evaluation | 教育测量与评估双语季刊 by an authorized editor of Chinese/English Journal of Educational Measurement and Evaluation | 教育测量与评估双语季刊.

The Development of Psychological and Educational Measurement in China

Zhang Houcan and Luo Fang

Beijing Normal University

1 Historical Highlights

In China today the ideas that “the human mind can be measured” and “testing is a useful tool to discriminate among minds” are well-accepted. These ideas can be traced back to the ancient Chinese philosopher and educator such as Confucius (551-479 B.C.), who classified people into three categories on the basis of intelligence: people of “great wisdom”, people of “average intelligence”, and people of “little intelligence.” His classification of people, to some degree, is equivalent to the nominal scale and the ordinal scale in the modern psychological and educational measurement.

Confucius also made personality evaluation of his students. For instance, he said, “*Student Ci is understanding, Qiu is artistic, Chai is clumsy, Shen is dull-witted, Shi is extreme, and You is reckless.*” In trying to use the evaluation as a guideline for personalized education, Confucius pointed out that “*As student Qiu is timid, we should give him a lot of encouragement; while student You is aggressive, discouragement is then needed.*” All these sayings clearly indicated his perceptive thinking about individual differences and individualized education.

One hundred and fifty years after Confucius, another great thinker, Mencius (327-289 B.C) became well-known for his saying, “*scaling makes it possible to understand weight, measuring makes it possible to understand length; these are true for all things, especially for the mind.*” This is a clear and important statement of the inevitability and possibility of quantitative measurement of the human mind.

Another root for the development of psychological measurement was the Civil Service Examination which originated in ancient China. The Civil Service Examination system was first established by the Chinese emperor to select personnel and to examine his officials’ fitness to their jobs. The contents of the exam underwent many changes. In the year of 1100 B.C. (Han dynasty), candidates were examined in the “Six Arts”: music, archery, horsemanship, writing, arithmetic and the rites and ceremonies of public and private life. The content of the Civil Service Examination got modified gradually and in the year of 200 A.D.

became “Five Studies”: civil law, military affairs, agriculture, revenue and the geography of the empire. After the seventh century, the Civil Service Examination system was developed into a national testing and selection system. It consisted of a variety of subjects and approaches for testing, emphasizing the ability to remember and interpret the Confucian classics. In addition to writing an essay on a given subject, participants took oral exams and composed poetry, conducted sentence completion extracted from classical writings, and made choices of synonyms andonyms. All these tasks and formats turned out to be quite similar to those used in modern testing for knowledge, logical and analytical thinking.

In the sixth century, another well-known scholar, Liu Xie, designed what seemed to be the first psychological test in the world. Scholar Liu’s test seems simple from today’s standards; for instance, he had people “*drawn a square with the left hand and a circle with the right hand at the same time, in order to demonstrate that under these conditions neither would be done correctly.*” However, this is exactly the “distraction test” developed by Binet in the 1890s.

From ancient books we know that there were some developmental tests for children around the sixth century in China. An article written by a scholar, Yan (531-590), indicated that the so-called “testing the child at age one” game was a popular custom in southern China.

On a child’s first birthday, he/she would be placed on a large table full of food, clothing, paper, pens, jewelry, toys, books, in addition to arrow and sword for boys, or needle and thread for girls. The baby was encouraged to crawl freely and pick up the item he or she liked best. By observing what the baby grasped first, the proud parents projected the baby’s intelligence, personality characteristics and aptitude.

This shows that people at that time had realized that there was a certain relationship between psychological characteristics and behavior, and this custom lasted until the 20th century.

Furthermore, there were some non-textual intellectual puzzles in ancient China, one of which was named “the Seven Boards of Cleverness” or Tangram. It was designed a thousand years ago during the Song dynasty (960-1277). It is composed of seven geometric shapes of different sizes, which can be arranged into a variety of designs. Solving the puzzle requires visual-spatial perception, divergent thinking and creativity, challenging the user’s potential for ingenuity. At the end of the nineteenth century, the puzzle was renamed “Intelligence-Benefit Graphs,” emphasizing the viewpoint that by training, intelligence, being a dynamic process, can be changed or improved. The Tangram was later developed into standardized paper-and-pencil tests that could be used to conveniently and accurately test groups of individuals. Another well-known ancient Chinese puzzle is the “Nine Connected Rings.” This puzzle consists of nine connected copper rings mounted on a bar with a rod running through the center of the rings. The goal of the puzzle is to get the rod out of the rings, which requires reasoning. These toys can be helpful for developing children’s intelligence, and are still quite popular in the world.

2 Modern Chinese Psychological and Educational Measurements

Chinese psychology and educational measurement have a long history, but scientific theories and methods of psychological and educational measurement came from the West. Early in 1915, Creighton used a Chinese translation version of psychological tests including mechanical memory, organized memory, alternation, metaphor, etc., to test 500 pupils to compare the intelligence of Chinese and American children. In the same period, Chinese scholars also began to design their own tests. Yu Ziyi, Director of the Jiangsu Provincial Teacher’s Affiliated Primary School, compiled the “Brush calligraphy Scale for Primary Students” in 1918 according to the design procedure of the Sandylke Calligraphy Scale. It was the earliest standardized educational test in China. In 1920, Liao Shicheng and Chen Heqin launched a measurement course at Nanjing Normal University and applied their own tests to evaluate students. Then the two co-published the book *Psychological Testing Method* and formally introduced the psychological measurement. The tests in the book point to the coordination of interest, observation, thought and action. It covers a wide range of topics, including graphics, mathematics, phrases and words, imagination and social knowledge, and morality. Among the tests, 23 of them are directly adopted from foreign test content, and 12 are based on the characteristics of Chi-

nese students (Ke, 2008). Liao Shicheng also enriched the test developed for primary and secondary school individuals and turned it into a group assessment, which was called the “Liao Group Test” (Yang, 2002).

In April 1922, the Binet-Simon scale was translated into Chinese by Jia Peijie and named as “Children’s Mind Development Measurement Method.” It was tested in some primary school students in Jiangsu and Zhejiang provinces. In the same year, Zhang Yaoxiang published an article on psychometrics and new law examinations in the *Education Series*, and for the first time, the psychological test was listed as one of the entrance examination subjects in Beijing Higher Normal School, proposed by Zhang. After 1922, articles about psychological and educational tests were published on the *Psychology* magazine edited by Zhang and other education magazines (H. Zhang, 2001), showing the prosperity of psychological and educational testing in the early 20th century in China.

At the same time, China Education Improvement Agency hired American expert W. A. McCall to help design various educational tests and train related personnel in China. After McCall came to China, he collaborated with education experts from all over the country to complete more than 50 tests, wrote the article *Scientific Measurement of Chinese Education*, and conducted two training sessions for graduate students (Gao, 1985). Led by the China Education Improvement Society, in 1923, general intellectual and educational tests were administered to primary school students across the country. Norms and other statistics were obtained for the age and class groups of students in grades three through eight. This large-scale investigation caught the attention of the education community at the time (Gao, 1985).

After 30 years of hard work starting from 1925, Ai Wei had compiled tests for all subjects in primary and secondary schools, and primary school children’s ability tests and intelligence tests, which indicated the beginning of test design and development in China. He designed eight tests for measuring reading comprehension (1935) and Chinese Language (1940), nine tests for measuring elementary school arithmetic application (1940) and high school plane geometry (1947), four tests for measuring middle school English (1943) and English Understanding Speed (1936), and nine tests for measuring junior high school common sense (1940). In 1946, two tests, the high school study ability test and high school and university scholastic test, were published. Ai also founded a pilot school in Nanjing in 1934; students were selected by their scores on these educational

tests and taught in accordance with their aptitude. In addition, Ai Wei and Zhang Yaoxiang designed literacy tests, respectively. With their colleagues, they conducted a survey of children's reading interests, compared results between reading aloud and reading in silence, and designed a reading test (Yang, 2002). Their efforts had greatly contributed to the development of education tests in China.

In 1931, the China Testing Society was established, marking a new era for the development of China's psychological and educational test. The China Testing Society aimed to study test theory, promote test methods, and cultivate test talents (Gu, 1991). It also founded the journal *Testing*, effectively promoting the development of testing at that time. In 1940, the China Testing Society membership grew to have 160 regular members. According to incomplete statistics, on the eve of the Sino-Japanese War, Chinese psychologists developed or adapted about 20 standardized intelligence tests and personality tests and more than 50 educational tests.

In the late 1930s, as the Sino-Japanese War began, the development of psychological and educational measurement, as well as most disciplines in China, was greatly affected and hindered.

After the July 7th Event in 1937, the journal *Test* was suspended. During the Sino-Japanese War, activities were not carried out in the name of the China Testing Society, but the testing continued, and the scope of the tests was gradually expanded. In addition, some educational psychologists used tests to conduct psychological research, such as teaching methods, Chinese characters, and national psychology, and so forth. These research all used tests and measurement as their research methods (Gao, 1985).

3 Chinese Psychological and Educational Measurement after the Founding of the People's Republic of China

3.1 Period of Stagnation

From 1949 to 1978, after the founding of the People's Republic of China, a movement to reform social sciences under the guiding principle of Marxism took place. Psychology, along with psychological and educational measurement, came under attack. Courses on psychology and educational measurement were no longer offered at academic institutions, and related research and applications were stopped. Not surprisingly, the psychological and educational measurement was considered to be *mentalistic* and *retroactive* during that time when individual differences

were denied. As a result, scientific research on psychological and educational measurement was totally abandoned.

After the launch of the Cultural Revolution in 1966, psychology was thoroughly criticized and even recognized as pseudoscience. Psychological measurement became one of the most damaged branches of psychology.

3.2 Period of Rebirth

In the late 1970s, the Cultural Revolution ended. People reacknowledged the role of psychological and educational measurement and various tests and recognized the real-world significance of understanding individual differences. Since then, psychological and educational tests had gradually restored their status and started to develop again in China.

Shortly after the resumption of psychology in 1979, at the Third National Academic Conference of the Chinese Psychological Society in Tianjin, two professors, Zhang Houcan and Hu Dehui, first called for the restoration of psychological measurement courses. Their suggestion received wide recognition and positive responses from a large number of participants. In response to the lack of psychological and educational measurement researchers, Lin Chuanding and Zhang Houcan initiated and taught, with Wu Tianmin, a national psychological measurement and statistics course in Wuhan in the spring of 1979. The offering of this course indicates a reemergence of psychological and educational testing in mainland China. Since then, China's psychological and educational measurement work has begun to recover and develop. The training course aimed to raise awareness of measurement, train talents and design materials for practice. In addition to teaching basic knowledge of statistics and fundamental principles of psychological and educational testing, Zhang Houcan and the other two professors also invited Wu Tianmin to share his experience of helping Lu Zhiwei revising the famous psychological measurement scale for children, the Binet scale. Lin Chuanding introduced the Wechsler Children's Intelligence Scale to the community and prepared for the future revision.

Since then, scientific research on psychological and educational measurement in China was reborn. In the autumn of 1980, academic institutions such as Beijing Normal University made lots of effort into psychological and educational measurement, laying an important foundation for this area to regain glory in China.

3.3 Period of Rapid Development

3.3.1 Theoretical Research

In the 1980s, well-equipped universities and colleges began to offer courses in psychological and educational measurement and actively trained professionals. Meanwhile, some researchers focused on promoting communications between Chinese and international communities, and introduced advanced educational theories and approaches from abroad, including a great number of textbooks and techniques. Through a series of academic presentations and articles, Professor Zhang Houcan from Beijing Normal University and Professors Qi Shuqing and Dai Haiqi from Jiangxi Normal University demonstrated the value of Classical Test Theory and Item Response Theory in psychometrics. The work of these precursors laid a solid theoretical and methodological basis for the development of Chinese psychological and educational measurement.

Subsequently, scholars from Beijing Normal University represented by Xin Tao and Bian Yufang initiated research on cognitive diagnosis evaluation and parameter estimation of measurement model. Later, Liu Hongyun conducted fruitful research on the tracking, intervention, and promoting of learning based on cognitive diagnosis evaluation.

At the same time, scholars represented by Dai Haiqi and Ding Shuliang from Jiangxi Normal University conducted many studies on the role of cognitive diagnosis in educational examinations, introduced the theory and methods of cognitive diagnosis test design, and discussed the application of multidimensional Item Response Theory and adaptive tests in psychological and educational tests.

Other universities gradually established and strengthened their measurement teams as well, including Nanjing Normal University, South China Normal University, East China Normal University, Huazhong Normal University, Northeast Normal University and Zhejiang Normal University. Researchers in these universities conducted various in-depth research on multidimensional Item Response Theory, testlet effect theory, and cognitive diagnose model of multi-dimensional testlet. Their work enriched the theoretical knowledge of Item Response Theory and cognitive diagnosis models, and facilitated the development of psychological and educational measurement in China.

The topic of test-score equating received much attention and has been thoroughly studies. For example, Liu Hongyun and Luo Fang (2010) in their study *Comparison of IRT True Score Equating and Observed Score Equating* explored two equating approaches, the IRT true score equat-

ing and observed score equating methods. Zhang Minqiang (2009) from South China Normal University in his study *Application Research of Test Equating In High School Entering Exam: An Example of High School Entering Exam of Foshan city in Guangdong Province* research how to select decent equating design and method to address the issue of converting scores of high school entering exams in different regions of China. In addition, studies on psychological and measurement has received the attention of researchers from universities other than the normal universities mentioned above. For example, Xie Xiaoqing from Beijing Language and Culture University studied Classical Test Theory and Item Response Theory-based equating designs and procedures and improved the framework of test score equating through a series of articles, including *Comparison of 15 Equating Methods* (2000).

Studies on Item Response Theory has spawned computerized adaptive testing (CAT), which has been widely used with the popularization and rapid development of computers. Since the publication of Qi Shuqing et al.'s (2000) *Designs and Applications of Computerized Adaptive Testing*, Chinese researchers have been conducting research in this area. Ding Shuliang and colleagues focused on the item selection strategy in CAT and published a series of articles, including *Item Selection Strategies of Computerized Adaptive Testing based on Graded Response Model* (Chen et al., 2006). These studies have advanced the development of CAT in China.

3.3.2 Test Design

China has adopted a strategy of starting from the revision of foreign tests and gradually developing tests in Chinese. This strategy has achieved good results in several years. The revised foreign tests include two categories, the Intelligence Test and the Personality Rating Scale. For intelligence tests, there are the third edition of the Internal Scale (Wu Tianmin), the second and fourth editions of the Wechsler Intelligence Scale for Children (WISC-R and WISC-IV), the fifth edition of which is currently being revised (Zhang Houcan), the Wechsler Intelligence Scale for Adults (WAIS) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Gang Yaoxian, Hunan Medical College), Raven's Progressive Matrices for group intelligence tests (H. Zhang & X. Wang, 1989), and the Snijders-Oomen Nonverbal Intelligence Tests (SON-tests; age 6-40; H. Zhang & J. Xu, 2012). For personality tests, there are the Minnesota Multiphasic Personality Inventory (MMPI; Institute of Psychol-

ogy, Chinese Academy of Sciences), Cattell's 16 Personality Factor (16PF; Li Shaoyi), Eysenck Personality Questionnaire (EPQ; Chen Zhonggeng), and the like. These tests are published in their Chinese versions, and have also been revised to establish a Chinese norm and widely applied in a long period of time, filling the research gaps and vacancies in China. Researchers in China have developed the Big Seven Personality Inventory of China, which corresponds to the Big Five Inventory, but the scale did not gain popularity.

Since the mid-1980s, due to the development of testing theory and technical knowledge, psychologists in China have begun to develop psychological testing tools suitable for Chinese. For example, Cha Zixiu from Institute of Psychology of the Chinese Academy of Sciences compiled the intelligence scale for identifying abnormal children. Wang Wenyun from Hangzhou University compiled a mental scale for screening mentally handicapped children. Other scholars also compiled a variety of scales for early childhood development. Zhang Houcan and her team published the CDCC Children's Developmental Scale of China (CDSC age 3-6). This CDSC scale is mainly used to examine the growth and intellectual development of three- to six-year-old children. It has a wide range of applications, including 16 sub-tests and four aspects of skills: language, cognition, social cognition, physical fitness and movement. In addition, Fan Cunren from Institute of Psychology of the Chinese Academy of Sciences compiled a child development scale for children below the age of three. These tests, compiled by domestic researchers, are closely linked to the Chinese culture and contemporary conditions. Thus, they are more suitable for practical applications, reflective of the development direction of Chinese psychological and educational tests, and are endorsed by users (H. Zhang, 2001).

3.3.3 Societies and Conferences

With the in-depth development of Chinese psychological and educational measurement, there is an increasing number of psychological and educational measurement professional organizations being established and related academic conferences being held.

At the Fifth National Academic Conference of the Chinese Psychological Society held in 1984, the Psychometric Professional Committee (now renamed as the Psychological Measurement Branch) was established as a sub-branch of the Chinese Psychological Society. This sub-branch consists of two groups: the psychological assessment group and the educational measurement group. The establishment of

this organization has played a positive role in promoting the development of psychological tests and in strengthening international academic exchanges.

In 1988, after the establishment of the Chinese Educational Statistics and Evaluation Society, high-level workshops were held several times. Some members of the branch have actively participated in the national college entrance examination standardization reform project. They have further contributed both to the planning and implementation of the standardization of the national college entrance exam, in terms of test design, item writing, scoring, and score report, and to the educational research and practice in their home provinces.

Regarding international knowledge exchange, starting from 1987 when Zhang Houcan gave a special report titled *Psychological Measurement in China* at the International Society for the Study of Behavioral Development (ISSBD) in Tokyo, Chinese scholars have delivered many studies on psychological and educational measurement at international academic conferences for many times. In 1990, China joined the International Test Commission (ITC), and Zhang Houcan was elected to its council. In 1991, from December 2nd to 5th, Nanjing Normal University held China's first international academic seminar on psychological testing.

In 1992, Mainland China began academic exchange activities with Chinese Taiwan. Under the leadership of Zhang Houcan, two academic organizations (two branches of the Chinese Psychometric Professional Committee — the Psychological assessment group, and the educational measurement and evaluation group) together with the Educational Statistics and Evaluation branch of the Chinese Society of Education established a stable relationship with Taiwan Society of Psychological and Educational Measurement through a biennial symposium. Since its first meeting in Taipei in December 1993, the Cross-Strait Psychological and Educational Measurement Symposium has been successfully held for 13 times in Taiwan and the mainland. More than 100 mainland scholars have visited Taiwan; nearly the same number of scholars from Taiwan have visited mainland China. Conference proceedings were also published after each meeting. This academic exchange activity not only expanded the sources of information for psychological and educational measurement researchers, but also promoted research enthusiasm. At the same time, it fostered mutual understanding, enhanced the friendship between scholars on both sides of the strait, and led to further collaborative research.

4 Psychological and Educational Measurement in the New Era

Since 2010, psychological and educational measurement has been more vigorously developed in China, which can be shown based on five aspects. First, **the number of researchers in the universities is increasing, whose research is also increasingly integrated with the international community.** A series of articles have been published in *Psychometrika*, *Applied Psychological Measurement*, *Educational and Psychological Measurement*, *Multivariate Behavioral Research* and other international mainstream journals for psychology and education measurement, such as Xin Tao et al.'s (2015) *Local Equating of Cognitively Diagnostic Modeled Observed Scores*, Chen Ping et al.'s (2016) *A new online calibration method for multidimensional computerized adaptive testing*, Zhen Chanjin et al.'s (2018) *The information product methods: A unified approach to dual-purpose computerized adaptive testing*, and Tu Dongbo et al.'s (2020) *Q-Matrix Estimation Methods for Cognitive Diagnosis Models: Based on Partial Known Q-Matrix*.

Second, **the National Assessment Center for Education Quality under Ministry of Education was established in 2007.** In July 2012, led by Beijing Normal University, the Collaborative Innovation Center of Assessment toward Basic Education Quality was established. Eight organizations joined the center as the core collaborative units, including East China Normal University, Central China Normal University, Northeast Normal University, Southwest University, Shaanxi Normal University, National Institute of Education Science, the National Education Examinations Authority, and iFLYTEK Co., Ltd.

The Collaborative Innovation Center of Assessment toward Basic Education Quality undertakes the task of assessing the quality of basic education in China. By integrating the professional strengths from related fields at home and abroad, the center builds a multi-disciplinary and multi-field basic education quality assessment team to develop multi-disciplinary and multi-field assessment index system and tools of basic education quality. The center also built a nationwide multi-level basic education quality assessment data collection network. In 2007-2014, the pilot assessment of basic education quality was carried out. From 2015, the quality assessment of basic education in China was officially carried out. Each year, two school subjects are assessed in the combination of language and art, mathematics and sports, science and moral education; every three years,

all the six subjects are assessed; and the national assessment finish one cycle every three years. The national assessment covers 31 provinces (municipalities, autonomous regions) and the Xinjiang Production and Construction Corps in Mainland China. In 2018, China's first *Compulsory Education Quality Assessment Report in China* was released. It provides an objective account of the students' moral, intellectual, physical, and artistic performance, as well as of the teaching and learning quality of schools in the compulsory education system in China. Further, it puts forward suggestions on how to further improve the quality of Chinese compulsory education system.

At the same time, provincial-level basic education quality assessment centers have been set up all over the country. Customized to local education characteristics, these centers carry out various activities on basic education quality assessment, identify issues in the basic education system, and ensure that more students can enjoy fair and quality education.

In 2015, the National Assessment Center for Education Quality under Ministry of Education and the Collaborative Innovation Center of Assessment Toward Basic Education Quality hosted the International Meeting of the Psychometric Society (IMPS) in Beijing. IMPS has a long history and is one of the largest scaled and the highest leveled international conferences in the field of psychometrics; the year of 2015 is the first time that IMPS was held in China since its establishment. There were 600 participants, including more than 200 scholars, researchers, and institutional personnel from overseas in the psychology and education field, as well as more than 300 Chinese scholars and researchers in research and assessment institutions. Experts and scholars presented the latest methods, technologies, and research progresses in the field, and shared their research ideas and results.

Third, **China actively participates in international large-scale educational assessment projects.** The OECD's Program for International Student Assessment (PISA) tests 15-year-old student's reading, mathematics, and science ability. The National Education Examinations Authority introduced and launched the PISA 2006 in China Research Project, conducted pilot tests in Tianjin, Beijing and Weifang, and obtained some valuable conclusions about the performance and education issues of Chinese 15-year-old students. On April 17, 2009, PISA 2009 was administered in 152 sample schools in Shanghai simultaneously, and 5115 15-year-old students participated in the assessment. This is the first time that Shanghai, representing

China, participated in the official PISA assessment. Subsequently, students from mainland China participated in PISA 2012, PISA 2015 and PISA 2018 assessments.

Fourth, **psychological and educational measurement researchers play a major role in the reform of the college entrance examination.** Concerning the whole society, the college entrance examination reform is a key area has a direct impact on all the other areas of the Chinese education system reform. On September 4, 2014, the national implementation of the examination and admission system reform was officially released. This is also the most comprehensive and systematic reform of the examination and the enrollment system since the resumption of the college entrance examination in the late 1970s.

This round of reform has made major adjustments in terms of the examination subjects and the admission mechanisms of colleges and universities. The reform of the enrollment system explicitly set up the standards that there will be no separation between liberal arts and sciences in the high school period, and the total score of the college entrance examination will consist of scores from both liberal arts and sciences. A multi-disciplinary admission mechanism based on the college entrance examination, high school academic level test scores, and comprehensive character evaluation was explored. The local education examination institutes around the country have also introduced local college entrance examination reform policies pertaining to local priorities.

Psychology and education measurement scholars not only undertake a lot of research work in the aspects of score equating, score reporting, test analysis, and so forth, but also play a major role in conducting core literacy assessment and high-order thinking ability assessment. For example, Xin Tao's (2019) *Exploring the Path of the Assessment of the New College Entrance Examination Goals* pointed out that establishing the standard references, promoting the comprehensive quality assessment into practice, and strengthening the basic education quality assessment system pave the way for achieving the goal of a healthy education system through the new college entrance examination. Also, improving testing processes and reducing differences among local assessments pave the way for achieving and maintaining social equity. Yang Xiaodong's (2020) *On the Misconceptions about Key Competences* discussed the concepts of key competences, highlighted the value of key competences on reforming the basic education curriculum, discussed misconceptions about key competences from the epistemology and conceptualization per-

spectives and emphasized the importance of connecting students' daily life with the social involvement.

Fifth, **Chinese psychological and educational measurement researchers keep up with the new wave of psychological and educational measurement research.** The rapid development of artificial intelligence and educational big data has brought opportunities for innovation in tools and methods for psychological and educational measurement. Chinese researchers have carried out numerous studies in the fields of automated scoring of essay, automated scoring of constructed response items, automated analysis of classroom images and videos, and new forms of assessments such as human-computer interactive assessment, game-based assessment, and procedural data mining. Published papers include but not limited to: Li Jian et al. (2015) proposed to assess psychological constructs through games. They modified tasks of online games and utilized the log-file of these tasks to measure participants' metacognitive planning. Liu Hongyun et al.'s (2018) *Analysis of Process Data of PISA 2012 Computer-Based Problem Solving: Application of the Modified Multilevel Mixture IRT model* utilized the process data recorded in one of complex problem-solving tasks in Program for International Student Assessment (PISA) 2012 and proposed a modified Multilevel Mixture IRT model (MMixIRT) to explore the problem-solving strategies. Luo Fang et al.'s (2019) *Automatic Generation of Number Series Reasoning Items of High Difficulty* proposed a new automatic generator for number series reasoning items for educational assessment, and developed a model for predicting item difficulties using item characteristics.

Zhu Tingshao from Institute of Psychology, Chinese Academy of Sciences, conducted a series of research on network-behavioral psychology to understand user personality, mental health, and social attitude based on user network behaviors, and consequently, to implement early warnings and effective intervention for at-risk groups. *Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study* (Cheng et al., 2017) employed computerized language analysis methods and developed a tool to automatically assess at-risk people's suicide risk levels and emotional distress in Chinese social media.

Luo Fang, Tian Wei, and other scholars of the Collaborative Innovation Center of Assessment toward Basic Education Quality have developed a computer-interactive testing tool for science subjects, which will be piloted in the 2020 national assessment of basic education quality.

5 Problems and Prospects

(1) Education reform in China has entered the deep-water area. Psychological and educational measurement researchers should actively shoulder the responsibility of education reform, participate in education reform, and use their professional knowledge and skills to deepen the reform of the college entrance examination and implement the core literacy reforms.

(2) The rapid development of artificial intelligence and educational big data has had a great impact on traditional psychological and educational measurement, including updating of traditional psychological and educational measurement methods, as well as assessment content. Educational measurement researchers should actively adapt to the development of artificial intelligence and educational big data, learn new theoretical models and research methods, and promote the development of research using educational big data in China.

(3) Enhance interdisciplinary and cross-disciplinary research capabilities. In the era of educational big data, research methods based solely on psychology and education are not sufficient for large-scale in-depth research. Psychological and educational measurement researchers should actively enhance their interdisciplinary and cross-disciplinary research capabilities, familiarize themselves with the latest developments in fields such as computer science, make use of emerging methods such as machine learning, deep learning, and learning analytics to conduct psychological and educational measurement research.

(4) Educational measurement researchers should vigorously strengthen international knowledge exchange and collaboration, actively participate in prominent international conferences, publish in renowned international journals, and jointly train high-quality psychology and education professionals with well-known foreign universities, and constantly absorb advanced theories and experiences of countries abroad. Meanwhile, Chinese psychological and educational measurement researchers should develop new assessments suited to conditions in China in the ever-changing era.

References

- Chen, P., Ding, S., Lin, H., & Zhou, J. (2006). Item Selection Strategies of Computerized Adaptive Testing based on Graded Response Model. *Acta Psychologica Sinica*, 38(03), 461–467.
- Chen, P., & Wang, C. (2016). A new online calibration method for multidimensional computerized adaptive testing. *Psychometrika*, 81, 674–701.
- Cheng, Q., Li, T. M., Kwok, C.-L., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. *Journal of Medical Internet Research*, 19(7), e243.
- Gao, J. (Ed.). (1985). *History of Chinese Psychology*. Beijing: People's Education Press.
- Gu, M. (1991). *Dictionary of Education: Volume 10. History of Modern and Contemporary Education in China*. Shanghai: Shanghai Education Press.
- Houcan, Z., & Xiaoping, W. (1989). Standardization Research on Raven's Standard Progressive Matrices in China. *Acta Psychologica Sinica*, 21(2), 113–121.
- Ke, X. (2008). *Biography of chen heqin*. Nanjing: Jiangsu Education Publishing House.
- Li, H., & Xin, T. (2019). Exploring the Path of the Assessment of the New College Entrance Examination Goals. *Journal of the Chinese Society of Education*, 40(11), 67–71.
- Li, J., Zhang, B., Du, H., Zhu, Z., & Li, Y. M. (2015). Metacognitive planning: Development and validation of an online measure. *Psychological Assessment*, 27(1), 260.
- Liu, H., Liu, Y., & Li, M. (2018). Analysis of Process Data of PISA 2012 Computer-Based Problem Solving: Application of the Modified Multilevel Mixture IRT model. *Frontiers in Psychology*, 9, 1372.
- Liu, Y., Luo, F., & Liu, H. (2010). Comparison of IRT True Score Equating and Observed Score Equating. *Journal of Psychological Science*, 33(3), 676–680.
- Qi, S. (2000). Designs and Applications of Computerized Adaptive Testing. *Jiangxi Educational Research*(2), 217–222.
- Sun, L., Liu, Y., & Luo, F. (2019). Automatic generation of number series reasoning items of high difficulty. *Frontiers in Psychology*, 10, 884.
- Wang, D., Cai, Y., & Tu, D. (2020). Q-Matrix Estimation Methods for Cognitive Diagnosis Models: Based on Partial Known Q-Matrix. *Multivariate Behavioral Research*, 1–13.
- Xie, X. (2000). Comparison of 15 Equating Methods. *Acta Psychologica Sinica*, 32(2), 217–223.
- Xin, T., & Zhang, J. (2015). Local equating of cognitively diagnostic modeled observed scores. *Applied Psychological Measurement*, 39(1), 44–61.
- Xu, J., Zhang, H., Du, Y., & Tellegen, P. J. (2012). A Re-

- view of Snijders-Oomen Nonverbal Intelligence Test. *Journal of Psychological Science*, 35(4), 984–987.
- Yang, X. (2002). *The History of Chinese Psychology*. Hefei: Anhui Education Press.
- Yang, X. (2020). On the Misconceptions about Key Competences[J]. *Journal of East China Normal University (Educational Sciences)*(10), 48–59.
- Zhang, H. (2001). The development of psychometrics in China. In C. P. Society (Ed.), *Contemporary Chinese Psychology* (pp. 209–214). Beijing: People's Education Press.
- Zhang, M., Li, G., & Jiao, C. (2009). Application Research of Test Equating In High School Entering Exam: An Example of High School Entering Exam of Foshan city in Guangdong Province. *Studies of Psychology and Behavior*, 7(1), 27–31.
- Zheng, C., He, G., & Gao, C. (2018). The information product methods: A unified approach to dual-purpose computerized adaptive testing. *Applied Psychological Measurement*, 42(4), 321–324.