Ministry of Education and Science of Ukraine

Mykola Gogol Nizhyn State University

Educational and Scientific Institute of Natural and Mathematical Sciences,

Medical and Biological Sciences and Information Technologies

Department of Information Technologies, Physical, Mathematical

and Economic Sciences

Specialty: 011 Educational measurements

**QUALIFICATION WORK**

to obtain a Master's degree

**The development and implementation of educational measurements in Chine**

**(history, modern state and perspectives)**

performed by student **Huang Libin**

**Supervisor:** **Zinchenko N.M.**,

Doctor of Physical and Mathematical Sciences, Professor

Admitted to the defense:  2024

Head of the department

Prof. Kazachkov I.V.

Nizhyn − 2024

ANNOTATION

The master's thesis “The development and implementation of educational measurements in China (history, current state and perspectives)” is devoted to the analysis of educational measurements in China.

The work consists of an introduction, the main part divided into four chapters, conclusions and a list of references. The first section contains a theoretical overview of the basic concepts of educational measurement, the concept of education quality and its structure. The second section provides a detailed overview of the history of educational measurement in China. The third section is devoted to a detailed description of the research sample, tools and processing of the results of international studies of the quality of national education systems: TIMSS, PISA, PIRLS and China's participation in these studies. The main results of the studies and their analysis are presented. The fourth section describes the main prospects for the development of educational measurement in China.

Keywords: educational measurement, quality of education, monitoring, international research.

**CONTENTS**

INTRODUCTION.....................................................................................................5

CHAPTER 1. ANALYSIS OF THEORETICAL AND METHODOLOGICAL FOUNDATIONS FOR THE IMPLEMENTATION OF EDUCATIONAL MEASUREMENT.....................................................................................................8

1.1. Content of the basic concepts of the theory of educational measurement.........8

1.2. Monitoring indicators and quality of education...............................................15

1.3. Testing method in monitoring studies..............................................................21

1.4. Basic information of international studies PISA, TIMSS, PIRLS and their tasks.........................................................................................................................23

CHAPTER 2. HISTORY OF EDUCATIONAL MEASUREMENT IN CHINA..28

2.1. Formation of the educational measurement system.........................................28

2.1.1. Early stages (before the XX century)............................................................28

2.1.2. Educational reforms in the early twentieth century.......................................31

2.2. Development in the second half of the XX century.........................................32

2.3. Educational Measurement in Modern China....................................................35

CHAPTER 3. CONTENT AND METHODOLOGY OF INTERNATIONAL STUDIES OF THE QUALITY OF NATIONAL EDUCATION SYSTEMS: TIMSS, PISA, PIRLS, CHINA'S PARTICIPATION AND ANALYSIS OF RESULTS................................................................................................................37

3.1. TIMSS - Trends in Mathematics and Science Education.................................37

3.1.1. Structure........................................................................................................37

3.1.2. Sampling........................................................................................................38

3.1.3. Toolkit...........................................................................................................39

3.1.4. Evaluation of results......................................................................................41

3.2. PISA - Program for International Student Assessment....................................41

3.2.1. Structure........................................................................................................41

3.2.2. Sampling........................................................................................................43

3.2.3. Toolkit...........................................................................................................43

3.2.4. Evaluation of results......................................................................................46

3.3. PIRLS - Program for International Reading and Levels Study........................47

3.3.1. Structure........................................................................................................47

3.3.2. Sampling........................................................................................................47

3.3.3. Toolkit...........................................................................................................48

3.3.4. Evaluation of results......................................................................................50

3.4. China in international studies of education quality..........................................50

CHAPTER 4. PROSPECTS FOR THE DEVELOPMENT OF EDUCATIONAL MEASUREMENT IN CHINA, CHALLENGES AND PROBLEMS....................58

4.1. Prospects for the development of educational measurement in China: innovations in educational measurement.................................................................58

4.2. Challenges and problems in educational measurement in China.....................60

4.3. Recommendations for improving the system of educational measurement.....61

CONCLUSIONS.....................................................................................................64

LIST OF REFERENCES........................................................................................66

**INTRODUCTION**

The development and implementation of educational measurement in China is a multifaceted process that spans several millennia and has a significant impact on global educational standards. The history of educational measurement in China dates back to the ancient dynasties, when the first systems of knowledge and skill assessment were established. Since then, China has come a long way from traditional assessment methods to modern approaches that integrate the latest technologies and methodologies.

In the current context, China is actively developing and improving educational measurement by introducing innovative approaches and international standards. This includes both adapting existing practices and creating new assessment methods that meet the needs of a rapidly changing educational environment. Considerable attention is paid to the quality of education and transparency of assessment, which allows to ensure a high level of student training.

Integration of international research has become an important aspect of the development of educational measurement in China. China actively cooperates with international organizations and institutions, such as the Organization for Economic Cooperation and Development (OECD) and UNESCO, to share experiences and implement best practices in its education system. In particular, participation in programs such as the Program for International Student Assessment (PISA) allows China to obtain valuable data to improve its approaches to assessment and improve the quality of education.

Prospects for the development of educational measurement in China include the further introduction of new technologies, such as artificial intelligence and big data analytics, to improve the accuracy and efficiency of assessment. An important part of this process is also the integration of educational measurement into a global context, which allows not only to improve the level of Chinese education but also to facilitate the exchange of experience and knowledge at the international level.

This paper aims to analyze the historical stages of the development of educational measurement in China, assess its current state, and identify the main directions and prospects for further development in this area.

The relevance of this study is that education plays a key role in the socio-economic development of any country. Studying the history, current state and prospects of educational measurement in China allows not only to understand the specifics of the Chinese education system, but also to identify effective practices that can be adapted in other countries. In addition, the introduction of modern technologies and international standards in educational measurement contributes to improving the quality of education and training, which is an important aspect in the context of globalization and rapid changes in the labor market.

The object of the study is the system of educational measurement in China. This includes the historical development, current state and prospects for the implementation of educational measurement.

The subject of the study is the development of educational measurement in China, the peculiarities of international monitoring studies and their impact on the quality of education.

The purpose of the study is to analyze the development and implementation of educational measurement in China, identify their key characteristics and impact on the quality of education, China's participation in international research in the field of education quality, and assess the prospects for their further development.

In accordance with the purpose, the following research objectives were set: to analyze the literature on the research topic, to reveal the content of the basic concepts of educational measurement, to consider the main stages of development of educational measurement in China, to study the organization and methodology of international research in the field of education quality, to analyze China's participation and results in these studies, to substantiate the prospects for the development of educational measurement in China.

Testing of research results - the main results of the study were presented at scientific and practical conferences and in publications:

- a publication on “China's experience of participation in international educational monitoring projects” in the collection of abstracts of the All-Ukrainian multidisciplinary scientific and practical conference “Youth and Modern Trends in Scientific Thought”.

- presentation and publication of the report “The development and implementation of educational measurements in China” at the XVI All-Ukrainian Student Scientific Conference “Prospects for the Development of Exact Sciences, Economics and Methods of Teaching Them”.

**CHAPTER 1. ANALYSIS OF THEORETICAL AND METHODOLOGICAL FOUNDATIONS FOR THE IMPLEMENTATION OF EDUCATIONAL MEASUREMENT**

**1.1. Content of the basic concepts of the theory of educational measurement**

In the context of the transformation of modern society, the problem of ensuring high quality of educational activities and the implementation of educational measurements, in particular, is the subject of careful attention of educational researchers. Educational measurement is a field in education that solves a set of problems on the border with other disciplines, including psychology, pedagogy, mathematics, statistics, computer science, and management theory.

The achievements in the field of educational measurement are applied in practice throughout the hierarchical system of relations in the educational sphere. The basis of this activity is to obtain, analyze and display sufficiently objective information regarding the successful development of the actual level of knowledge, skills and abilities acquired by students. Reliable information will help to optimize management activities, with the global goal of radically changing the real situation for the better, as well as effective decision-making by educational management.

Leading scholars in the psychological and pedagogical literature note that there is a set of factors that determine the priority and relevance of educational measurement in modern education, namely

- a fundamental change in the educational paradigm and, in this regard, the need for an in-depth analysis of innovation processes in the hierarchical education system;

- development and implementation of innovative methods and forms of training and informatization at all levels of the educational hierarchy, as well as development of mechanisms for forming managerial influence on educational processes with simultaneous systematization and specification of evaluation actions;

- diversity of pedagogical technologies, variability of approaches to the organization of training, individualization of education, including the growth of exclusive educational programs, creation of effective means of controlling the quality of educational services and knowledge assessment in accordance with the current state standards in the field of education;

- internationalization of education, which ensures the creation of a single European and global educational environment, preserving the unity of international standards and national traditional requirements for the quality of education.

The field of educational measurement has evolved over several stages. The beginning of educational measurement dates back to ancient China. Back then, this country used a special procedure for competitive selection for vacant positions in government agencies. It was then that the foundations were laid for assessing the business qualities and knowledge and skills of civil servants. At the next stage (the Middle Ages), a point system for assessing real achievements during training was developed and applied in certain countries. However, the system itself has not survived to this day, and only some references in the relevant literature remain.

At the end of the nineteenth century, simple tests were used to test students' knowledge, particularly in writing and arithmetic. This was the first primitive way of assessing students' abilities. Over time, the development of the theoretical foundations of pedagogy, educational psychology, sociology and a number of other scientific disciplines gave rise to the development and justification of not only quantitative but also qualitative assessments, which were determined according to the degree of manifestation of a particular characteristic or property. As is well known, qualitative assessments are less informative and unambiguous than quantitative ones. That is why attention has recently been paid to the development of the latter.

The foundations of the modern theory of educational measurement were laid in the 1980s. In accordance with the new understanding of the content and structure of measurements, they are considered as a formulation of a certain function that more or less objectively reflects the empirical structure through the use of a congruent numerical structure.

In order to analyze the relevant theoretical foundations of educational measurement, it is necessary to clarify the essence and meaning of such concepts as "control", "measurement" and "evaluation".

In particular, the concept of "pedagogical control" can be considered as a single didactic and methodological system of control over the quality of educational processes. For a permanent educational process, such control is aimed at separating and adequately assessing the results of students' learning activities. In reality, despite the existing approaches and generally accepted criteria for deciding on the appropriate assessment in a certain point system, the assessment itself is still quite subjective.

In the process of pedagogical control, there is actually no object of measurement, no reference level (a certain zero), and no units of measurement that quantitatively reflect quality. This leads to relevant difficulties and a number of problems that teachers try to solve in practice. In doing so, they rely on the so-called mathematical theory of measurement.

The problem of ensuring a certain degree of objectivity in measuring the results of certain educational activities, according to leading scholars and teachers, is as follows. Measurement should be realized through an objective quantitative comparison of the assessed knowledge, skills and abilities of students with a certain standard (which can be realized in accordance with one of the conditional models). In addition, in the practical implementation of pedagogical measurement, knowledge, skills and abilities play the role of the property being assessed. And the virtualization of the standard for such measurement is realized by formulating control tasks or individual components of such tasks in the subject [2].

The concept of "measurement" is usually interpreted as a certain means for determining the parameters of the relevant object, or an algorithm, as a result of which the numerical value of the characteristic to be determined is compared with some standard that is absolutely equivalent to the selected measurement unit.

Measurement procedure is a set of relevant operations that allows you to move away from the concept of the so-called empirical tasks to the corresponding numerical values of the set of parameters that are determined (measured).

Measurement is both a process of "transferring" the parameters of an object to be determined to a certain numerical axis (or plane), and also an assessment, i.e., in fact, a procedure (algorithm) implemented using scales of individual personal abilities, in particular, in order to quantify the intellectual (or other) behavior of an object in addition to qualitative assessment. In turn, a scale is a certain numerical system in which the relationship between different parameters of the objects, processes and phenomena under study are reflected in the properties of some other set, which is usually a set of numbers [4].

Educational measurements provide an opportunity to assess the level of knowledge of students and the quality of the educational process in an educational institution.

The development of the very concept of "measurement" and the problem of measurement determines the solution of three interrelated tasks: why, what and with what to measure?

The answer to the first question is related to the purpose of measurement. And if the goal is to determine the quality of students' training in the course of certification, then the main focus is on identifying the amount of knowledge or skills acquired and their compliance with the norms (requirements of state educational standards). The answer to the second question is determined by the definition of the subject of measurement, and the answer to the third question is determined by the tools used in the course of measurement. If the method of measurement is testing, then it is considered as a process of measuring quantitative indicators using a test.

The quality of measurement results is assessed in terms of their validity. Validity, in turn, reflects the adequacy of empirical results to the global goal of the measurement. The analysis of the validity of the results should be systematic and multidimensional.

A measurement scale is a certain numerical system in which the relationship between individual parameters and characteristics of the phenomena under study, as well as the relevant processes, are reflected in the characteristics of a certain set, usually relative numerical indicators [1].

There are several types of scales, among which the most commonly used are:

- nominal scale;

- ordinal or ranked scale;

- interval scale;

- ratio scale.

1. A naming scale, or nominal scale, is formed by assigning "names" to objects. The naming is based on the comparison operation, which is primary for the construction of any scale. Objects are compared with each other, and their equivalence or non-equivalence is determined. As a result of this procedure, a set of equivalence classes is formed - objects belonging to one class are equivalent to each other and different from objects belonging to other classes. Equivalent objects are assigned the same names. Thus, all objects are divided into non-intersecting subsets and distributed to the classification cells.

This is the simplest measuring scale. It only classifies by name. In the name scale, only mutually unambiguous (identical) transformations are allowed. The purpose of using this scale is to classify objects. In this scale, numbers, if used, are used only as labels to distinguish objects.

2. The ordinal scale, or order scale, is more complex than the name scale. It does not classify according to the principle of "equivalent - non-equivalent", but according to the principle of "more - less". If in the scale of names it was indifferent in what order the classification categories (classes) were arranged, then in the ordinal scale they form a sequence from the category of "least value" to the category of "greatest value" (or vice versa).

The fundamental difference between the order scale and the naming scale is that the order scale organizes objects by a certain attribute. This introduces the most important concept - the property being measured. A transitional variant of the naming scale to the ordinal scale is the dichotomous classification: 1 - "there is a property", 0 - "no property".

An important aspect is the number of classes in the ordinal scale. By definition, there should be at least three classes in an ordinal scale, for example, "positive reaction - neutral reaction - negative reaction". However, it is irrational to place the entire variety of objects in only three classes, because objects that are quite different from each other can fall into the same class. In addition, the more classes there are in the scale, the more opportunities there are for testing statistical hypotheses (the greater the resolution of statistical criteria). On the other hand, if the number of classes is equal to the number of objects, as in a forced ranking, there is a danger of artificially exaggerating the differences between objects.

The ordinal scale and the name scale are the main scales of qualitative attributes and are referred to as non-metric scales. In many specific fields of science and practice, the results of qualitative analysis can be considered as measurements on these scales.

Scales of quantitative attributes (metric scales, or higher-type scales) are mainly interval scales and ratio scales

3. The interval scale, or interval scale, classifies objects according to the principle "more by a certain number of units - less by a certain number of units". In this scale, the numbers reflect not only the difference between objects in terms of the degree of expression of a feature, but also how much more or less this feature is expressed. An important feature of this scale that distinguishes it from another metric scale, the ratio scale, is the absence of a natural reference point and a natural unit of measurement. The researcher must independently determine the starting point and choose the unit of measurement. The main limitation of the interval scale is that it does not allow you to determine how many times one object is larger than another in terms of the value of the measured property.

4. Ratio scales are the most common. They have a natural starting point - zero, which means the absence of a value. In such a scale, numbers reflect not only the difference between objects in terms of the degree of manifestation of a feature and how much more or less the feature is expressed, but also how many times more or less it is expressed. This is an even more flexible scale, where, in addition to defining equality, rank order, and equality of intervals, equality of relations is also known. This scale establishes the equality of the relationship of numbers attributed to objects.

Measurements in education are used to assess the achievements of students in educational institutions of various levels and types, including students of higher education institutions. They can be used to quantify the level of knowledge, skills, and abilities of individuals participating in testing, for example, during school exams. Moreover, they are often used to assess the abilities of applicants for a vacant job, etc.

It should be emphasized that measuring the quality of education based on its compliance with certain standards is possible only if such standards themselves are of sufficient quality [1].

As noted by E. Bennett (USA) at the conference of the International Association for Educational Measurement held in October 2013, since the world is changing rapidly, educational measurements should meet certain principles: provide a significant amount of information; meet a large set of goals; be based on modern scientific ideas about competence as the basis for test design; balance test designs and test tasks; use assessment and scoring of results in accordance with modern scientific achievements; adapt modern methods

In the field of educational measurement, empirical research methods are used: observation, questionnaires, interviews, testing, analysis, scaling, experimentation. However, the most commonly used is testing. It performs the following functions: diagnostic, control, training, educational, motivational, developmental, and prognostic.

Thus, educational measurements, as the highest manifestation of evaluation, should be included in the structure of the system of monitoring the quality of education in educational institutions of all levels. They form the basis of the scientific and pedagogical justification of the process of obtaining the most objective, both quantitative and qualitative assessments in the diagnosis of the educational process. The role of a measuring device in pedagogical measurements is most often played by tests. The most effective use of the test system in the educational process should be not only as a means of diagnosing and monitoring the learning achievements of students, but also as a means of teaching. It is necessary to skillfully use both educational, developmental and motivational functions of test control. Such a systematic approach should ensure the quality of the educational process at all its stages.

# 1 .2. Monitoring indicators and quality of education

The concept of "monitoring" was first used in environmental sciences for continuous observation in order to control, predict development and protect the environment. Usually, a distinction is made between global, regional, and local environmental monitoring, the main function of which is to study the state and predict the development of the environmental object under observation.

In the most general terms, monitoring can be defined as the continuous observation of a process in order to determine its compliance with the desired result or initial state.

A monitoring object is a system that is targeted by monitoring procedures. The main feature of monitoring objects is dynamism [5]. All objects studied with the use of information monitoring are constantly changing, developing, and are subject to external influences that may cause undesirable changes in the object's functioning.

The subject of information monitoring is the state of the system under study in certain periods of time and specific changes within this system.

Monitoring subjects are the bearers of monitoring functions who perform these functions. They are conditionally divided into two large groups: entities that provide information and entities that collect and process information. In other words, the subjects in the monitoring system are both institutions, structures and individuals.

A set of monitoring indicators is a set of primary and secondary indicators that can provide a holistic view of the state of the system, qualitative and quantitative changes in it.

The monitoring toolkit includes a set of statistical reporting forms, information standards, questionnaires, questionnaires, etc.

Monitoring tools are various technical means used by monitoring entities in their activities.

Monitoring is considered in connection with the quality of education. Today, the scientific literature is not clear on the definition of the concept of quality of education. In the general context, it is understood as:

* compliance with the standard and norm;
* Possession of the characteristics of full compliance with the standard and the levels of achievement of the desired results;
* the result and quality of the educational process, the quality and value of the means of achieving the goals, the quality of the environment;
* physical, spiritual and psychological health, general culture, intelligence, and value orientations;
* general knowledge of the effectiveness of a particular educational system;
* a set of important properties and characteristics of educational outcomes that can satisfy its consumers.

There are several hundred definitions of the concept of quality of education, depending on the content that is included in the understanding of the essence of education, what tasks the education system should solve at each stage of its development. The most common opinion is that the quality of education is compliance with a certain norm or standard.

Depending on the meaning of the term "quality of education", its essence is considered as:

* learning outcomes and the educational process;
* efficiency of the educational institution;
* the quality of the education system at a certain level;
* a certain desirable ideal of a person's education;
* is a priority of the state educational policy [19].

The quality of education is characterized by multidimensionality, multidimensionality and multivariate.

Thus, the quality of education is the resulting, integral function of all these factors. It should be noted that until recently, pedagogy considered it in the context of substantiating the system of knowledge quality by assessing the results of the educational material learned by students, their ability to operate with knowledge and apply it in specific situations.

The problem of measuring and evaluating the quality of education, determining the list of its characteristic features, searching for quality indicators and criteria is being solved by scientists from all over the world within the framework of the science called qualitology.

Qualitology is a threefold science that encompasses:

* Quality System theory;
* theory of quality assessment (qualimetry - Assessment, Evaluation);
* the theory of quality management (Management and Monitoring of Quality).

Each of these three components has a certain set of criteria and indicators of education quality that make it possible to comprehensively assess any education system both in terms of its external and internal parameters and performance results. Not only does a person's own success as an individual, but also the development of the entire state, its competitiveness and prestige on the world stage depend on how freely he or she can navigate modern life, adapt to different types of activities and conditions, and learn new professions and assimilate huge amounts of information.

Qualimetry is one of the components of qualitology. Qualimetry is a science that develops ways to measure the qualitative characteristics of any object or system.

Qualimetrics of education addresses the issue of finding methods and technologies for measuring the qualitative characteristics of the educational system in general or the education system in particular. In the practice of measuring and evaluating the quality of education, the effectiveness of an educational institution, an educational system of a certain level, and the effectiveness of teacher teaching, so-called qualimetric models are used [18].

Qualimetric models used in the educational sector are analytical models of a certain object of the education system (educational institution, department or education administration, teacher's activity, school head), which are built on the basis of the method of expert assessments using a qualimetric approach to determining the content and weight of each quality indicator of the object of assessment.

Qualimetry as a component of the general theory of quality assessment involves:

* defining a system of criteria, indicators, and indicators of the quality of education;
* Selecting assessment methods that are appropriate to the instrument (i.e., the measurement tool) and measurement procedures;
* Conducting monitoring studies to manage the quality of education at different levels of educational systems (from national to institutional)

Summarizing various approaches to ***defining the essence of education quality, we*** can interpret this pedagogical category as a certain set of properties and characteristics of the educational process that give it the ability to form a level of general and professional competence that will meet the needs of citizens, enterprises and organizations, society and the state. It is an indicator of the development of society in a certain time dimension, so it should be considered in the dynamics of its changes in relation to the factors that determine its nature. In a broad sense, the quality of education is understood as a balanced compliance of the process, the result and the educational system itself with the goals, needs and social norms (standards) of education.

Thus, the quality of education can be represented as a multidimensional model of social norms and requirements for the individual, the educational environment in which his or her development takes place, and the education system that implements them at certain stages of human learning. It is a multifaceted category that inherently reflects various aspects of the educational process - philosophical, social, pedagogical, political, demographic, and economic.

Naturally, the problem of education quality is primarily a pedagogical one: it is aimed at improving the result in specific learning conditions by pedagogical means. However, this problem cannot be limited to the pedagogical aspect, as it also requires a managerial solution. In this case, it acquires a managerial meaning, moves into the realm of educational management and becomes a problem of education quality management. Moreover, monitoring as a systemic procedure is not a pedagogical technology. This procedure is mainly a component of management activities. The problem of education quality is complex. Therefore, it cannot be solved only within the framework of pedagogical theory and educational practice. One could even say that it is more managerial, although one should not go to the other extreme and limit oneself to this aspect.

The quality of education can be interpreted as the degree of satisfaction of the participants of the educational process with the educational services provided by the educational institution or the degree of achievement of the educational goals and objectives. Also, the quality of education can be seen as compliance with a certain norm or standard. In other words, quality determines the usefulness, value of objects, their ability to meet certain needs or realize certain goals, norms, i.e. expresses adequacy to requirements, needs, and norms. It is known that the concept of norm is not absolute, unchanging, fixed. Therefore, the category of "quality of education" is revised and changes depending on the conditions, requirements of society and time.

The structure of education quality:

* quality of the educational process as a result of pedagogical activity;
* quality of educational programs, educational literature, textbooks and manuals;
* quality of professional training and qualification of teaching and research staff;
* quality of resources and the learning environment in which the educational process takes place (legal, financial, personnel, scientific, methodological, material and technical);
* the quality of students' personal traits and abilities;
* quality of state and public management of the education system;
* the effectiveness and diversity of the institution of external evaluation of the quality of education (national system of monitoring the quality of education);
* quality of conducting and interpreting the results of monitoring studies in the education system;
* quality and efficiency of public education administration.

Objectives of education quality assessment:

* improvement of pedagogical tools;
* determining the effectiveness of teaching and upbringing of students;
* comparison of educational institutions, rating;
* determining the efficiency of funds and resources;
* planning and forecasting the development of the educational sector in a particular region;
* formation of the educational policy of the state and the region;
* determining the prestige and competitiveness of the national education system.

In order to assess educational systems as a special indicator of social development, countries conduct special international studies aimed at determining the level of education achieved by a country based on certain indicators. As a rule, they evaluate national education systems in a ranking ratio and do not aim to compare them or determine the best ones.

Thus, monitoring in education is a system of measures for collecting, processing, analyzing and disseminating information in order to study and assess the state of functioning of a particular educational entity or the educational system as a whole and to predict their development based on the analysis of the data obtained and the identification of trends and patterns.

**1.3. Testing method in monitoring studies**

Testing is a method of psychological diagnostics based on the use of standardized questions and tasks (test items) that have a certain scale of values. It is used for standardized measurement of individual differences of a person (object - pupil, student). There are three areas of testing application:

* education in connection with the increase in the duration of training and the complexity of curricula;
* professional training and selection due to increased growth rates and production complexity;
* psychological counseling - in connection with the acceleration of sociodynamic processes.

Tests are characterized by their versatility and high objectivity. Pedagogical testing is used to measure certain personal qualities, the level of knowledge and skills during training, professional selection, etc.

The main features of a truly well-organized test process are the scientific nature, efficiency, objectivity and quality of the test results obtained. The issues of test creation, analysis and interpretation of measurement results are within the competence of ***testology*** - the science of test creation and use. In the field of pedagogical measurements, it is the theoretical, methodological and methodological substantiation of the processes of test development and application.

Testing is one of the methods of pedagogical assessment that can be carried out for different purposes and, accordingly, be current and final. For each case, a special test should be developed that meets the purpose and objectives of the test.

A test is a testing tool and must meet a number of conditions that essentially make it a measurement tool, not just a simple set of questions and tasks. Any test consists of test items that can be presented in different forms and formats and for which certain test item quality characteristics are calculated.

A test item is a component unit of a test that meets the requirements of manufacturability, form, content, and statistical requirements, namely: known difficulty, sufficient variation in test scores, and positive correlation of test scores with scores across the entire test.

Standardized tests are tests that have undergone a special methodological experiment to develop norms (criterion-referenced scoring systems), test reliability and validity on the basis of a representative sample, and have a well-described instructional and methodological apparatus. The standardization process is very long (up to 10 years). Such a test can be used, for example, to conduct a final certification exam in strict compliance with the requirements of the testing technology. If the test standardization procedure is violated at least partially, this leads to a loss of the test's measuring properties, limits the possibilities of its application and dramatically increases the errors in the test results, sometimes even eliminates the possibility of interpreting the results for analysis and drawing reliable conclusions. The use of testing as a method of obtaining reliable and objective information about certain properties of the object under study requires the creation of a so-called standardized test.

Requirements for tests as measurement tools:

* the test must be valid;
* the test must have high accuracy (and therefore a small measurement error);
* A test must be reliable, meaning that the results obtained with its help must be reproducible and their values stable. The validity of a test ensures its reliability, but not vice versa.
* the presence of a scale that provides a certain level of measurement. According to the degree of increase in the level of measurement, scales are divided into: nominal, ordinal (rank), interval and ratio scales;
* the test must be standardized and certified. This requirement is met in the case of final certification testing. If the testing is carried out during the educational process, i.e. is ongoing, this requirement is not mandatory [39]. The development of testing and the use of tests in different spheres of human activity, for different purposes, was marked by the emergence of a variety of tests. The question of the need to classify tests has become natural, but despite the development of different approaches to classifying tests on different grounds (purpose, form, content, etc.), there is no universally recognized classification yet.

**1.4. Basic information of international studies PISA, TIMSS, PIRLS and their tasks**

Among the many UN development goals that China supports, the global goal is that quality education should be accessible to all. The international community recognizes that the main factors of human quality of life are well-being, education and health, and the main goal and priority of the development of societies in the 21st century should be the quality of education.

The educational policy of many countries is based on a quality strategy. China is no exception. The quality of education is placed on a par with such concepts as accessibility and efficiency. The development programs of the Chinese educational system are focused on achieving these goals. Almost all documents of the Chinese government, on which China's educational policy is based, emphasize the development of world-class quality education with Chinese characteristics. The material, financial, human and scientific resources of society and the state are directed to its provision. Therefore, one of the main objectives of China's education policy is high-quality education at all stages and levels, performance assessment and quality management.

Global comparison of educational outcomes, launched by transnational organizations such as the Organization for Economic Cooperation and Development (OECD) and used to assess the quality of education in different countries. This trend is becoming increasingly common and involves the use of educational inspection to ensure better results on national and international tests. Transnational, large-scale assessments such as the Program for International Student Assessment (PISA), the Triennial International Mathematics and Science Study (TIMSS), and the international Project for International Reading and Literacy Study (PIRLS) have had a variety of impacts on countries in their education reform and quality assurance efforts, including prompting them to review their policies and practices regarding school inspection. When implementing new reforms, Chinese policymakers are constantly looking for new ideas outside the country, which are then adapted to Chinese realities. The search for better systems to assess the quality of education in China is an ongoing process. It is the participation of many countries in international programs such as PISA, TIMSS, or PIRLS that has aroused deep interest on the part of the Chinese central government, which is trying to keep up with global trends.

The OECD is actively implementing the Programme for International Student Assessment (PISA), which aims to determine whether students who have completed compulsory education (at the age of 15) have the knowledge and skills to function fully in society. It provides information on students' performance in reading, mathematics and science literacy and identifies factors that influence the development of these skills at school and in the family. It is expected that the information will enable participating countries to make informed decisions in the education sector.

The age of 15 was chosen because in many countries, compulsory schooling ends by this age, and curricula in different countries have much in common. At this stage, it is of particular interest to determine the state of knowledge and skills that may be useful for students in the future, as well as the ability to independently acquire the knowledge necessary for successful adaptation in the modern world.

From 4500 to 10000 students in each country are tested.

The results of the study are:

* quantitative indicators characterizing the state of basic knowledge and skills of 15-year-old pupils;
* Quantitative indicators that characterize the state of factors that affect student learning outcomes and the state of affairs at the school;
* Quantitative indicators that characterize trends in the results over time.

The study of pupils' training is conducted in three areas: "reading literacy", "math literacy" and "science literacy".

Each of the areas corresponds to specific school subjects. Particular attention is paid to students' understanding of the basic concepts, mastery of the basic methods studied in the three areas, and the ability to use their knowledge in different situations. It does not measure the level of mastery of specific subject content, but rather the level of broader knowledge and skills needed for adult life that are acquired through the study of school subjects.

Since the PISA study aims to test the availability of life skills and the readiness of young people for "adult" life, this approach distinguishes it from other international studies, the main purpose of which is to test the subject knowledge and skills defined by school programs by performing learning tasks that have little or no connection to real life [25].

The TIMSS project, an international study to assess the quality of mathematics and science education, is organized as a cross-section, but differs from others in that the cross-sections of the educational system were conducted not in one grade, but in two other consecutive grades: 3-4 and 7-8.

According to D. Wylit and R. Wolf, the developers of the project concept, the approach used in TIMSS is a compromise between conventional cross-sectional and longitudinal studies. It provides much more information about student achievement and the factors that influence learning outcomes. In a conventional cross-sectional study, the research is focused only on a specific age or year of study [24].

The purpose of the TIMSS study was to compare the science and mathematics proficiency of secondary school students in countries with different educational systems and to identify factors that influence proficiency.

TIMSS tasks were implemented in the following areas:

* collecting, analyzing, and summarizing information on educational systems in the countries participating in the study;
* analysis and comparison of science and mathematics education systems at the level of defining goals, planning educational content and requirements for students' readiness within specific disciplines (analysis of curricula, programs and textbooks of science and mathematics disciplines);
* analysis of the educational process in mathematics and science (organization of the educational process; teaching methods; opportunities provided to students in the study of mathematics and science; provision of the educational process with literature and other teaching aids);
* Assessment of learning outcomes, which involves not only assessing students' academic achievements in science and mathematics education, but also identifying their attitudes toward learning;
* analysis of the relationship between planned educational goals and levels of implementation, on the one hand, and learning outcomes (achieved level of education), on the other.

The results of the TIMSS study make it possible to compare countries by the dynamics of changes in the level of readiness of students during the transition from primary to secondary school.

The international project "PIRLS" - "Project on International Reading and Literacy Study" - is a monitoring study of the quality of education.

The purpose of the study is to:

* Comparison of the level and quality of reading and comprehension of text by primary school students in different countries;
* identifying and interpreting differences in national educational systems to improve the process of learning to read;
* identifying factors that influence educational outcomes [26].

The study was organized by the International Association for the Evaluation of Educational Achievement (IEA) in 2001. 40 countries participated in the project. PIRLS is scheduled to be conducted every five years.

In each school, students in the final grade of primary school were tested, students and their parents were surveyed, and teachers and school administrators were interviewed. The study assessed two types of reading that students most often engage in during school hours and outside of school:

* reading for the sake of gaining literary experience;
* reading for the purpose of learning and using information.

According to the types of reading identified, texts were selected and tasks for them were developed. The following skills were assessed while reading different types of texts:

* finding information specified in a certain form;
* formulation of conclusions;
* interpreting and summarizing information;
* Analyzing and evaluating the content, language, and elements of a text.

**CHAPTER 2. The History of Educational Measurement in China**

**2.1. Establishing a system of educational measurement**

**2.1.1. Early stages (before the XX century)**

Educational measurement in China has a long history, combining deep traditions with the evolution of modern approaches. This chapter examines the main stages of educational measurement development, its impact on society, and its transformation in response to changing political, social, and economic conditions.

The philosopher Confucius (551-479 B.C.) classified people according to their intelligence: "Sons of Heaven" (those with innate wisdom), "Noble Men" (those with knowledge), and "Blacks" (people who are unable to comprehend knowledge). His classification is to some extent equivalent to the nominal and ordinal scale in modern psychological and pedagogical measurement.

Another famous thinker, Mencius (327 - 289 BC), argued that it is inevitable and possible to quantify the human mind.

The first mention of educational measurement in China dates back to ancient times. Already in the Han Dynasty (206 BC - 220 AD), educational institutions began using exams to assess students' knowledge. This approach became the basis for the further development of the education system in China. During this period, the first measurement systems used for education and research were created.

The famous scientist Liu Xie developed the world's first psychological test back in the sixth century. From ancient books, we also know that there were tests for child development, the so-called "one-year-old child test" game. People at that time understood that there was a relationship between the behavior and psychological characteristics of a child.

The educational system of traditional China has deep historical roots, and a key component of it for over 1300 years was the imperial examination system known as keju. This system was launched in the 7th century during the Sui Dynasty (581-618) and was widely developed during the Tang Dynasty (618-907).

"Keju has existed for more than 1300 years, forming a unique approach to educational assessments based on standardized written exams.

The purpose of the keju was to ensure the meritocratic selection of officials for public office by assessing their knowledge of classical Confucian texts. Candidates passed a series of written exams that assessed not only their erudition but also their ability to analyze philosophical texts and formulate their own thoughts. The system's peculiarity was its relative openness: in theory, anyone, regardless of social status, could pass the exams and obtain high positions. This made keju a unique phenomenon for its time.

However, the system also had significant drawbacks. Much attention was paid to mechanical memorization of texts, and the range of disciplines was limited. Nevertheless, keju influenced the development of the idea of standardized testing in China and the world. The criticism was that the keju system focused on memorizing texts (especially Confucian canons), which limited the development of creative thinking.

In 1905, the system was abolished due to its inefficiency in the context of the country's modernization.

During the Song Dynasty (960-1279), many reforms were introduced to the education system, including the standardization of measurements. Public schools were established, allowing more people to be educated in core subjects such as literature, philosophy, and the arts.

The examination system was expanded, allowing ambitious individuals to gain positions in the government based on their knowledge and skills. This contributed to the formation of a class of apprentice officials who played an important role in government. The use of new teaching methods and technologies, such as printed texts, increased access to knowledge. There was also an increase in the role of teachers. Teachers received more support and recognition for their work, which contributed to the quality of education.

These reforms have significantly improved the quality of education and contributed to the development of cultural and intellectual life in China.

During this period, Tangram or, as it is also called, the Seven Boards of Wit was developed. With it, a child learns to analyze images, identify geometric shapes in them, visually break a whole object into parts, and vice versa - to make a given model from elements, and most importantly - to think logically.

The Ming Dynasty (1368-1644) was an important period in Chinese history, and during this time there were significant changes in educational dimensions. Here are some of the key aspects.

During the Ming Dynasty, more accurate methods of assessing students' knowledge were introduced. Measurements became more standardized, which ensured a level playing field for all students and increased the objectivity of assessment.

New teaching tools and methods were created. For example, new textbooks and teaching materials were developed to help students learn better.

New technologies and teaching methods have been introduced, making the learning process more efficient and understandable for students.

The Ming dynasty actively supported the development of science and research. Numerous academies and research institutes were founded, which contributed to the improvement of education and research.

Educational measurements have become a tool for assessing scientific achievements and the level of training of future scientists.

The keju exams continued to play an important role in the education system. They became more complex and demanding, which helped to select the most talented and prepared candidates for public office.

The reforms aimed to make the exams not only about memorizing texts, but also about the ability to analyze and think logically.

Education reforms during the Ming Dynasty had a significant impact on the development of educational measurement in China. The introduction of more precise measurement methods and the creation of new tools for learning contributed to improving the quality of education and ensuring the training of qualified specialists. These changes laid the foundation for the further development of the educational system in China.

During the Qing Dynasty (1644-1912), several important reforms in education were carried out. Here are some of them:

1. The foundation of Tongwen Guan (1861): A school was founded to teach European languages, mathematics, astronomy, and chemistry.

2. Western missionaries promoted new approaches to teaching and assessment. For example, St. John's University in Shanghai, founded in 1879, used written tests similar to those used in the United States.

3. The founding of Peking University (1898): The university was founded with a course based on the Japanese education system.

4. Abolition of the Imperial Examinations (1905): Imperial examinations were abolished and traditional academies were converted into Western schools. Various textbooks were also published throughout the country.

**2.1.2. Educational reforms in the early twentieth century**

The Chinese government began to introduce elements of the Western educational system. In 1904, a new school system was introduced that included regular assessments of students' knowledge.

Exams began to focus on scientific and technical disciplines to meet the needs of industrialization.

These reforms have contributed to the modernization of the educational process in China and prepared the country for the challenges of our time.

The end of the nineteenth century was marked by a crisis in China's traditional education system. Amid internal political turmoil and external pressure, reformist forces in the country demanded change. In 1905, the Qing Dynasty abolished the keju, which was an important step toward modernizing the educational system.

Modernization included borrowing educational measurement methods from other countries, especially Japan, Germany, and the United States. Important changes included:

-Transition from Confucian content to the study of natural sciences, foreign languages, and mathematics.

-The introduction of new forms of written exams.

-Development of the first standards for secondary and higher education.

During this period, China began to build the foundations of a modern educational system, with an emphasis on scientific approaches and practical knowledge.

Scientific theories and methods of psychological and educational measurement came from the West. In 1915, Creighton used tests translated into Chinese to test 500 students and compare the intelligence of American and Chinese students. During this period, Chinese scientists began to develop their own tests. The very first standardized educational test was created in China in 1918. In 1920, Liao Shicheng and Chen Heqin launched a measurement course at a teacher training university and used their own tests to assess students. Later, they published the book Psychological Testing Method and officially introduced psychological measurement. Liao Shicheng supplemented the test that was developed for elementary and middle school students and turned it into a group assessment, which was called the "Liao Group Test".

In 1922, for the first time, a psychological test was included among the subjects of the entrance exams in Beijing.

At the same time, the China Education Improvement Agency has hired an American expert to help develop educational tests and train relevant personnel in China.

In 1931, the Chinese Testing Society was established to study the theory of testing and promote testing methods. This society founded the Testing magazine, which effectively contributed to the development of testing.

**2.2. Development in the second half of the XX century**

After the establishment of the People's Republic of China (PRC) in 1949, the educational system was significantly reformed. The new government sought to unify the educational system in accordance with socialist ideals. Educational measurement became an important tool for realizing these changes. The main goal was to ensure that education was accessible to all citizens and that socialist ideals were introduced into the educational process. This meant that education had to contribute to the development of a socialist society by providing all citizens with the opportunity to receive education and develop their knowledge.

In 1952, a system of nationwide university entrance examinations, known as the Gaokao, was introduced. This exam has become a key tool for determining who is eligible for higher education. The Gaokao is held annually and includes various subjects such as Chinese, mathematics, a foreign language (usually English), and additional subjects depending on the student's choice. The results of this exam determine which universities and colleges may be available to candidates.

These reforms have made China's educational system more unified and accessible, and have contributed to the development of socialist ideals in society.

The Cultural Revolution in China (1966-1976) had a significant impact on education and brought about significant changes. During this period, the educational system suffered significant losses due to political ideologies.

In 1966, during the Cultural Revolution, the university entrance exam (Gaokao) was abolished. Instead, political and social factors, such as family background and participation in revolutionary movements, influenced admission.

Assessment during this period was based on the study of ideological texts, such as the works of Mao Zedong.

Educational dimensions were significantly reduced due to political ideology. The educational system was transformed to serve the ideological goals of the Communist Party. Educational programs were simplified to meet ideological needs, and many traditional subjects were removed from the curriculum.

During the Cultural Revolution, many universities were closed or transformed into other institutions. This led to a significant reduction in higher education opportunities. Standards of educational measurement dropped significantly. All assessments were replaced by ideological criteria

There was a decline in quality standards of education. Curricula were simplified, and teachers were often replaced with politically sympathetic ones, even if they were not sufficiently qualified.

These measures had a lasting impact on the Chinese educational system, and many of them were gradually changed after the end of the Cultural Revolution.

In the late 1970s, a transition to rapid economic development took place, and China's education system was reoriented to meet the needs of modern industries. The general trend in Chinese education of "fewer students and higher educational standards" dramatically raised the prestige of the education system. By the end of the 20th century, the network of universities grew rapidly, the government allowed the opening of private educational institutions and began decentralizing the general management of education.

In the late 1970s, the cultural revolution came to an end. The role of psychological and educational measurement and various tests was again recognized and since then, psychoeducational tests have gradually regained their status and started to develop again.

After the end of the Cultural Revolution, Deng Xiaoping initiated reforms aimed at restoring academic standards, returning to the idea of quality education and systematic educational measurement. In 1977, the Gaokao was reopened, and more than 5 million students took the exam, competing for 270,000 university places.

In addition, China has begun to actively integrate modern international approaches to educational measurement.

In the 1980s, the country returned to systematic educational measurement and assessment of student learning. At this time, well-equipped universities and colleges began to offer courses in psychological and educational measurement and to actively train specialists.

China has begun to focus on international assessment systems such as PISA and TIMSS to improve the quality of education.

**2.3 Educational Measurement in Modern China**

In the 2000s, the government introduced reforms aimed at integrating critical thinking and creative tasks into the assessment system.

Pilot projects. In provinces such as Shanghai and Guangdong, experiments were conducted to simplify test tasks and introduce project work to assess creativity.

In 2010, Shanghai scored highly in PISA thanks to reforms in the testing system that emphasized practical knowledge.

At the beginning of the 21st century, China is actively developing its educational measurement system. The country participates in international studies, such as PISA (Program for International Student Assessment), which allows to assess the level of knowledge of students in comparison with other countries. The use of standardized tests and modern methods of knowledge assessment helps to improve the quality of education and ensure its compliance with international standards [22].

The search for better systems to assess the quality of education in China is an ongoing process. Over the past decade, Chinese education inspectorates have undergone significant changes in the scope and focus of their reforms. In China, the State Education Inspectorate has been used by the government as an administrative tool for managing and controlling education, similar to the educational inspectorates used in other countries. The educational inspection is called jiaoyududao ("educational supervision") in Chinese and plays the role of management, supervision and consultation in the field of education. The Institute of Education Inspection sets regulations for local governments on how to run schools in accordance with policies and laws. However, the participation of a large number of countries in international programs such as PISA or TIMSS (International Mathematics and Science Studies) has aroused deep interest on the part of the Chinese central government, which does not want to lag behind global trends In 2010, a draft document "China's Long-Term Program for Education Reform and Development for the Period 2010-2020" was released, calling for the modernization of the Chinese education system at all levels to create a working quality-oriented management system and redistribute educational resources among schools to improve the quality of education.

In April 2015 The Steering Committee on Education of the State Council of the People's Republic of China proclaimed the National Compulsory Education Quality Monitoring Program, which introduces annual mandatory monitoring of the quality of education throughout the country. This event marked the creation of the Chinese National System of Compulsory Education Quality Monitoring. In the same year, the Ministry of Education of China established a specialized Center for Monitoring the Quality of Basic Education, designed to assess the academic achievements of students in all provinces of China once a year.

The main body of the educational inspection system in China is the Bureau of Educational Inspection, which develops regulations and standards for the supervision of education at the national level. It performs supervisory, evaluation, inspection, and monitoring functions to ensure the quality of education at all levels in accordance with the law. Inspectors prepare reports on the state of national education, monitor the implementation of national legal and policy requirements, and provide feedback to local education authorities. They also assess the quality of educational institutions. In particular, the School Supervision Authority is responsible for school inspections.

Thus, the history of educational measurement in China demonstrates a gradual transition from traditional assessment methods to modern standardized tests and international research. The implementation of educational reforms at different stages of history has allowed China to achieve high results in education and ensure its high quality.

The historical development of educational measurement in China reflects a complex interplay between tradition and innovation. From imperial examinations to modern standardized tests, the system has evolved to meet political, economic, and social demands. China's current success in international rankings is largely based on the lessons of the past.

**CHAPTER 3.** CONTENT AND METHODOLOGY OF INTERNATIONAL STUDIES OF THE QUALITY OF NATIONAL EDUCATION SYSTEMS: TIMSS, PISA, PIRLS, CHINA'S PARTICIPATION AND ANALYSIS OF RESULTS

**3.1. TIMSS - Trends in International Mathematics and Science Study**

**3.1.1. Structure**

TIMSS is organized by the International Association for the Evaluation of Educational Achievement (IEA). It is an independent international organization that represents a community of national research organizations and government agencies, known for its comparative research since 1959. Many research centers and professional organizations around the world participated in the study and development of its tools: ETS - Educational Testing Service, USA, Statistics Canada, IEA Secretariat, Netherlands, IEA DPC - IEA Data Processing Center, Germany and others. The entire study was coordinated by ISC - International Study Center at Boston College, USA. The structure diagram can be seen in Fig. 1 [24].

IEA

Statistics Canada

IEA DPC

Data Processing Center

IEA Secretariat

ETS

Educational Testing

ISC

International Study Center

TIMSS

Figure 1. The structure of TIMMS

Through comparative studies of the IEA:

* provides international benchmarks by identifying strengths and weaknesses in the education system;
* provides high quality data that points to key factors that influence teaching and learning;
* ensures high quality data to support educational reforms;
* promotes research in educational assessment and the development of global research in this important area.

**3.1.2. Sampling**

A very important criterion in the sampling process is to select countries that have already participated in previous TIMSS studies to allow for the identification of trends in math and science education.

Let's look at the selection structure. First, absolutely all schools in the participating countries are considered, regardless of the age of the students. This is an internationally desirable target group. Schools are excluded from it, depending on their geographical location or language of instruction. In some countries, for example, schools were selected that teach in the official language. This is a national exclusion from the sample [24].

At the next stage, the state is involved in the selection of schools. This is a nationally desired target group. The selection is based on the following school factors: small number of students, geographical remoteness, inconsistency of the curriculum with the state curriculum, teaching children with disabilities. These are outliers at the school level.

Next, after passing the previous stages, the national target group is determined. Students within the school are excluded from this group based on the following factors: students with intellectual disabilities, students with physical disabilities, and students who do not speak their native language.

As a result of all the stages, the final effective target group is determined [24].

**3.1.3. Toolkit**

TIMSS is one of the longest-running surveys, which is both an advantage and a disadvantage. On the one hand, participating countries have the opportunity to track their own trends across a wide range of indicators, while on the other hand, the developers are forced to retain much of what was stated in the first cycles without changing the basic approaches.

The TIMSS Assessment Frameworks and Specifications is used as the basis for developing the TIMSS research tools, which defines general approaches to assessing educational achievement in mathematics and science; development of tests and test tasks; describes the content of mathematics and science studied; lists the main factors that characterize students, teachers and educational institutions; for the analysis of which, information is collected in the course of the survey.

The research tools include:

* achievement tests;
* questionnaires (for students, teachers, school administrators, education experts, and research quality observers);
* methodological support (guidelines for national coordinators on organizing and conducting the survey, guidelines for sampling, guidelines for school coordinators, guidelines for testing, guidelines for checking open-ended questions, guidelines for data entry, etc;)
* software (for selecting classes and students, for data entry).

In developing the instrument, it is first and foremost recognized that TIMSS is inextricably linked to the content under study. Therefore, the subject content was selected and formalized at the intersection of two types of selection requirements: relevance to regular school content and relevance to mathematics and science education in general. Or in other words: what is actually taught and what should be taught.

Tests are developed based on the following principles:

* adequate coverage of the content and types of educational and cognitive activities studied;
* maximum relevance of the test content to the material being studied in most participating countries;
* Ensuring the communication of tests of all cycles;
* the importance of the content studied in terms of the development of mathematics and science education;
* Relevance to the age characteristics of the students for whom the test was developed;
* Compliance with the requirements for mass studies.

Since the study was planned as a long-term one (and initially it included elements of a longitudinal study), the most important requirement for the instrument was to ensure that tests from different years were linked and met the requirements for mass monitoring studies.

TIMSS includes two types of tasks: multiple-choice tasks, where the student selects the correct answer from four answer options, and open-ended or extended response tasks, where the student must provide a written answer:

* multiple-choice items provide a valid, reliable, and cost-effective measure of a wide range of content in a fairly short test time;
* open-ended or extended response tasks allow students to give an explanation of the answer: building diagrams, graphs, tables, explaining cause and effect relationships.

Potential micro- and macro-level factors were also identified - school and home environments, family environment, socioeconomic factors of the family and school as a whole, as well as features of training programs, curricula, age of students at the time of testing, and much more.

In fact, at the stage of identifying and defining these factors, the basic hypotheses of the study were formulated: the factors that influence academic performance in the subject areas under study were hypothetically formulated. When selecting the areas to be studied, they are differentiated into potentially related and unrelated.

**3.1.4. Evaluation of results**

The results are evaluated by professionals specially trained for this purpose. All potentially possible types and forms of responses are coded in a standardized manner. When checking, each answer is appropriately categorized and marked with the appropriate code. Thus, in common databases, it is possible to find out not only the number of points received by a given respondent, but also, for example, to see the tasks that he or she did not start.

Based on the results of the tasks, after the appropriate processing procedures, the entire set of respondents' answers is divided into four levels:

- advanced level of math skills (625 points or more),

- high level of math skills (550-624 points),

- average level of math skills (475-549 points),

- low level of math skills (400-474 points).

It should also be borne in mind that if a student's math proficiency score is below the lowest level, it should not be concluded that this student cannot perform any math activities. It is simply that he or she was not able to successfully apply his or her mathematical knowledge in most of the tasks offered in the study.

**3.2. PISA - International Program for International Student Assessment**

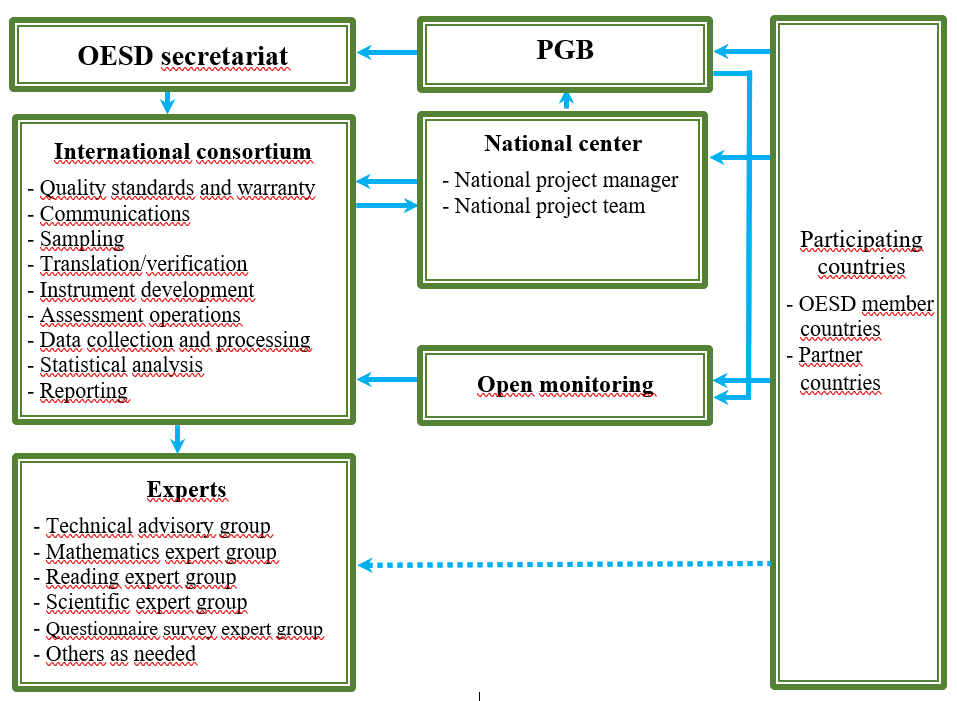
**3.2.1. Structure**

The Program for International Student Assessment (PISA) was developed and implemented under the auspices of the OECD - Organization for Economic Cooperation and Development.

The PISA project has several structural levels. The project is overseen by the OECD Secretariat, which is located in Paris. The PISA Governing Board (PGB), which consists of delegates from the OECD and participating countries, determines the project's policy. Members of the PGB meet twice a year. It can be said that PISA is based on the idea of close international cooperation in education. Each country establishes a PISA National Center and appoints a National Project Manager (NPM) whose responsibility is to coordinate all project-related activities within the country. The NPM usually works closely with the country's Chairperson to develop a forward-looking plan for policy decisions related to the project, conduct in-depth analysis, and report on country-specific results [8].

The National Center has a small team of experts working on the project, which is responsible for project design, implementation and reporting at the national level. This group works closely with the international staff who implement the project. One of the most important tasks of local experts is to analyze the tasks proposed for the participating countries. These tasks must be culturally appropriate and in line with the country's curriculum. All tasks and questionnaires used in the study have high psychometric characteristics and are valid. The structure is shown in Fig. 2.

Figure 2. The structure of PISA



The program is implemented by a consortium of leading international research organizations, with the participation of national centers and the OECD organization. The consortium is led by The Australian Council for Educational Research (ACER). The consortium also includes the following organizations: Netherlands National Institute for Educational Measurement - CITO; US Educational Testing Service - ETS; National Institute for Educational Research in Japan - NIER; WESTAT USA [25].

Today, PISA is the most comprehensive and rigorous international study to assess general academic achievement, as well as the breadth and coverage of additional (contextual) information about students, their families, institutional factors, and characteristics of the education system. The totality of these data allows us to identify the most significant factors and trends in education on an international scale.

**3.2.2. Sampling**

The construction of the sample of countries is quite simple: division into OECD member countries and partner countries.

Methodologically, PISA is conducted on a representative sample of countries. For example, in PISA 2006, about 400,000 15-year-old students participated, representing about 20 million children of this age studying in the 57 countries studied. And in PISA 2022, approximately 690,000 students from 81 countries and economies took part. This study covered about 29 million 15-year-old students. Each student completed written tasks for two hours [39].

**3.2.3. Toolkit**

The data collection procedure involves students and school principals answering questionnaires. These questionnaires are extremely important for finding out the characteristics of the students and schools participating in the testing. All of them are available on the program's website. The purpose of the questionnaires is to collect information about:

* students, their families, including their economic, social and cultural assets;
* students' attitudes toward learning, their habits at school and outside of school, and the atmosphere at home;
* quality of the school's material and intangible resources, public or private control and funding of the school, decision-making process, specifics of school staff selection, priority areas in the curriculum, availability of clubs;
* school structure and type, number and size of classes, school and classroom microclimate, and different types of reading-related activities in the classroom;
* peculiarities of the reading learning process, including students' interest, motivation, and activity in the classroom.

Participating countries are also offered 3 more questionnaires to choose from:

* a questionnaire on participants' awareness of computer technologies, which aims to collect information about the participants' ability to use a computer and their attitude to computer technologies;
* a questionnaire that collects information about gaps in the learning process, changes in the school system, expected learning outcomes, and tutoring outside of school;
* a questionnaire for parents that focuses on parents' own experiences with reading, availability of books at home, and parental involvement in their child's school life.

Data from mandatory and additional surveys are collected and analyzed. This allows us to develop indicators that describe the overall structure of the education system and the impact of these indicators on the country's economic system.

Reading is a complex activity, and not all of its elements are measurable. For the PISA test, the following three components of reading are recognized as the most important:

* situations - various reading goals and contexts in which to navigate with the help of the text;
* text - various reading materials;
* reading skills - cognitive strategies and ways of working with the text.

In order to create tools for these three components of reading activity, they need to be operationalized: defined through specific characteristics that allow test developers to assign test materials and tasks to certain categories. The same categories are used to analyze test results. It is clear that reading is a holistic activity, and its individual components do not exist in isolation - each in its own clearly defined department. The division of texts and questions into separate categories adopted in the PISA study does not imply a strict division of these categories. The classifications presented below are intended to create a balanced test and to provide a framework for describing test results.

The PISA test is designed to provide information about the reading literacy of students who are completing their compulsory education. Each test item models the conditions of real-life reading tasks that adolescents and adults face in and out of school. In terms of difficulty, the test items range from the simplest, requiring a superficial and literal understanding, to the most complex, requiring a deep, sophisticated, multi-level understanding. Test creators can vary the difficulty of a task based on the following ideas about the factors that determine this difficulty.

In order to track trends in the development of reading literacy, the PISA test maintains a general proportion of open and closed questions from year to year. Open-ended questions involve more or less detailed free responses that require expert judgment. Students' answers to closed questions do not require expert evaluation, they can be evaluated in an automated mode. The most classic closed questions involve choosing one correct answer from four or five options given in a ready-made form. More complex types of closed questions involve, for example, the location of the story's events in a time sequence (put the right numbers in a given list of events).

The same category of questions that do not require expert evaluation also includes those where the reader has to write one or two words, for example, write out a word from the text (while there is only one word in the text that is the correct answer to the question).

It should be emphasized that for comprehending and evaluating the message of a text, the open-ended form of the question is essential, since it is necessary to evaluate, first of all, the quality of the reader's reasoning, not the "correctness" of his or her answers [8].

**3.2.4. Evaluation of results**

The study uses a 1000-point scale. The international 1000-point scale had the following characteristics: the mean value was 500 points, the standard deviation was 100, which meant that about 2/3 of all students participating in the country study had results in the range of 400 to 600 points.

This means that for each task (based on the results obtained), its level of difficulty is calculated in points. For the set of tasks solved by the test subject, he also received a score (on the same scale).

As noted in one of the reports of the Center for Educational Quality Assessment, and this should be noted, the analysis of the results of individual countries and the construction of the international scale took into account the specifics of the tasks in all countries. If contradictory data were obtained for a number of tasks in individual countries, such tasks were excluded from the international analysis for all countries or only for one or more countries. Items with errors, such as typing, were also excluded from the analysis. For example, out of a bank of 150 reading tasks, 141 tasks were used in the construction of the international scale (i.e., 9 tasks were excluded from the final calculations).

Definition of the concept of "literacy" in relation to the main subject areas studied in PISA:

Reading literacy is the ability of a person to understand, use, reflect on, and interact with written texts to achieve their goals, develop their knowledge and opportunities to participate in society.

Mathematical literacy is the ability of a person to identify and understand the role that mathematics plays in the world, to make informed judgments, and to use mathematics to meet the needs that ensure the life of a constructive, engaged, and thinking citizen.

Scientific literacy is scientific knowledge and its use to ask reasonable questions, to acquire new knowledge, to explain scientific phenomena, to draw data-based conclusions about phenomena; understanding of the main characteristics of science as a form of human knowledge, the role of science and technology in changing the material, intellectual and cultural environment; interest in scientific explanation of the world, characteristic of a thinking citizen.

**3.3. PIRLS - Reading and Intermediate Reading Levels Study**

**3.3.1. Structure**

The international study Progress in International Reading Literacy Study PIRLS is organized by the International Association for the Evaluation of Educational Achievement (IEA). This organization was established more than 50 years ago, and since then it has established itself as an influential international structure, providing a variety of studies in different subjects (mathematics, science, technology, and reading). In particular, it is important to emphasize that the PIRLS study is conducted by the same international organization that conducts the TIMSS study. The structure is shown in Fig. 3.

**3.3.2. Sampling**

The sample is constructed similarly to the TIMSS sample. The international PIRLS survey is conducted every five years. It was first conducted in 2001. Note that 35 countries participated in the 2001 PIRLS  study, 41 countries in the 2006 study, 50 countries and 9 cities from the benchmarking group in 2011, 48 countries and 8 cities in 2016, and 57 countries and 8 benchmarking participants in 2021.

Preparations for the next cycle of the study begin three years before the date of the official study.

IEA

Statistics Canada

IEA DPC

Data Processing Center

IEA Secretariat

ETS

Educational Testing

ISC

International Study Center

PIRLS

Figure 3. The structure of PIRLS

**3.3.3 Toolkit**

The PIRLS survey instrument was developed in 2001 and has been improved since then, but in such a way that the 2006 cycle and subsequent ones will be the basis for valid comparisons.

PIRLS was organized to meet the highest quality standards. The participating countries had to meet high sampling requirements to prevent errors and ensure comparability of results.

According to the research concept, two types of reading are evaluated, which are most often used by students during classes and outside of school:

* reading for the purpose of acquiring literary reading experience;
* reading for the purpose of learning and using information [10].

Since the PIRLS study is a monitoring study, and one of its main tasks is to study the dynamics of reading acquisition, a necessary condition for conducting the study was the reuse of some of the texts and tasks for them in each cycle.

The undoubted advantage of the PIRLS study is a well-thought-out system of evaluation of the work performed, combining qualitative and quantitative assessment. Multiple-choice and sequence-of-events tasks were always assessed with one point; open-ended tasks were assessed with 1 to 3 points depending on the complexity of the task. The tasks are checked by experts in accordance with the international task assessment document, which contains both general approaches to assessing the performance of each task and specific examples of student responses that are assigned a particular score.

To collect information on factors that allow for the interpretation of the study results, the PIRLS toolkit included, in addition to texts and tasks, the following:

* a questionnaire for the student;
* a questionnaire for teachers;
* a questionnaire for parents;
* a questionnaire for the school administration.

In 2006, the PIRLS research project became one of the most representative studies of primary schooling. It involved 215,000 students from 40 countries.

In each school, students in the final grade of primary school were tested, students and their parents were surveyed, and teachers and school administrators were interviewed. The study assessed two types of reading that students most often engage in during school hours and outside of school:

* reading for the sake of gaining literary experience;
* reading for the purpose of learning and using information.

According to the types of reading identified, texts were selected and tasks for them were developed. The following skills were assessed while reading different types of texts:

* finding information specified in a certain form;
* formulation of conclusions;
* interpreting and summarizing information;
* analysis and evaluation of the content, language, and elements of the text [26].

The International Reading Test was developed by the world's leading experts in the field of reading testing and teaching. All test versions consist of two parts. One part contains a literary text and the other a non-fiction text. All test items have the same structure and level of difficulty. In nine test variants, the texts and tasks are printed in one workbook. In one variant, Nature, the texts and tasks are presented in different notebooks: the texts are in a reading book and the tasks are in a workbook. The test takes more than two hours (135 to 155 minutes) [30].

The PIRLS study can be seen as complementary to the PISA study, as it assesses the reading achievements of primary school graduates.

**3.3.4 Evaluation of results**

The rating scale in the study ranges from 0 to 1000. The scores are divided into levels that reflect different degrees of reading competence.

The main levels of assessment in PIRLS:

* Beginner level (less than 400 points): Students can solve only the simplest tasks that require basic reading skills.
* Low level (400 to 474 points): Students are able to read simple texts and perform basic comprehension tasks.
* Intermediate level (475 to 549 points): Students can understand and interpret more complex texts and complete tasks that require deeper analysis.
* High level (550 to 624 points): Students demonstrate the ability to think critically, understand complex texts and can apply their knowledge in different contexts.
* Advanced level (625 points or more): Students have a high level of reading competence, including the ability to analyze, synthesize and critically evaluate information from texts.

This division helps to determine what level of reading skills students are at and what areas need further development.

### 3.4. China in international studies of education quality

China has been participating in PISA since 2000. It was represented by the provinces of Beijing, Shanghai, Jiangsu, and Guangdong (B-S-J-G-China). Since 2009, Shanghai has been represented in this study. Subsequently, Beijing, Jiangsu and Guangdong joined in 2015, and Jiangsu and Zhejiang in 2018 and 2022. This allowed China to present a broader picture of its education system.

By participating in PISA, China not only demonstrates its achievements, but also gets the opportunity to compare its educational system with other countries, which helps to identify areas for further improvement.

With a population of over 20 million people, Shanghai is one of the largest cities in the world. While Beijing serves as China's political capital, Shanghai is the business capital and most international city. Although Shanghai is home to only 1% of China's population and occupies 0.6% of its territory, the city generates 1/8 of the total national income. The city has long enjoyed a special status in China, allowing it to implement certain public policy strategies, including education, at its own discretion, while other cities and provinces are required to strictly follow the policies set by the national government in Beijing. Shanghai was one of the first cities to be granted the right to independently approve university entrance exams and reform the curriculum.

In 2015, China's participation in PISA was extended to Beijing, Jiangsu, and Guangdong.

The capital city of Beijing, home to more than 20 million people, has invested heavily in primary and secondary education, with the largest allocation to education compared to other provinces and municipalities. The student-teacher ratio in Beijing is well below the national average. Beijing continues to struggle with inequality of access to education and uneven school quality. Similar to Shanghai, Beijing has implemented a number of reforms, including curriculum reforms to make learning more engaging and integrated, and reform of the university entrance exam. The government also seeks to improve the quality of all compulsory schools to reduce competition for places in the "best" schools.

Jiangsu is a province on the east coast of China, located north of Shanghai, with a population of nearly 75 million people. The province began reforming its school system back in 1993 in an effort to equalize resources between urban and rural areas and expand extracurricular opportunities in high schools. Today, 686 training bases have been established in Jiangsu to provide high school students and adults with practical skills in specific industries.

Guangdong is a southern coastal province that is the largest and richest in China. It is considered an economic center focused on advanced manufacturing. However, there are significant gaps between urban and rural areas. In 2017, the World Bank launched the Guangdong Compulsory Education Project, funded by a $120 million loan, to improve school facilities and teacher training in rural areas of the province.

In PISA 2015, the results of four provinces are presented as "BSJG (China)". These regions are treated as a single entity and the reports are not broken down by province or municipality. The total population in these four provinces is 230 million (out of China's 1.357 billion total population). With the participation of these provinces in the study, more rural students were able to take part in PISA, compared to previous years when only Shanghai represented the country. The results of the four provinces were more modest than in 2009 and 2012, when Shanghai topped the rankings by a wide margin. The BSJG China provinces ranked 10th in science, 24th in reading, and 6th in math.

In the PISA 2022 study, China demonstrated strong results. It is particularly worth noting that Macau (China) was ranked second in the math ranking, after Singapore. In addition, Hong Kong (China) also showed good results.

Let's take a closer look at the results on the example of Hong Kong.

Hong Kong first joined PISA in 2002. By comparing results internationally, Hong Kong policymakers and educators can analyze the policies and practices of other countries.

Average results in 2022 were lower than in 2018 in math and reading, and about the same as in 2018 in science. After achieving the highest results in 2012, there was a decline in performance in all three subjects between 2012 and 2022.

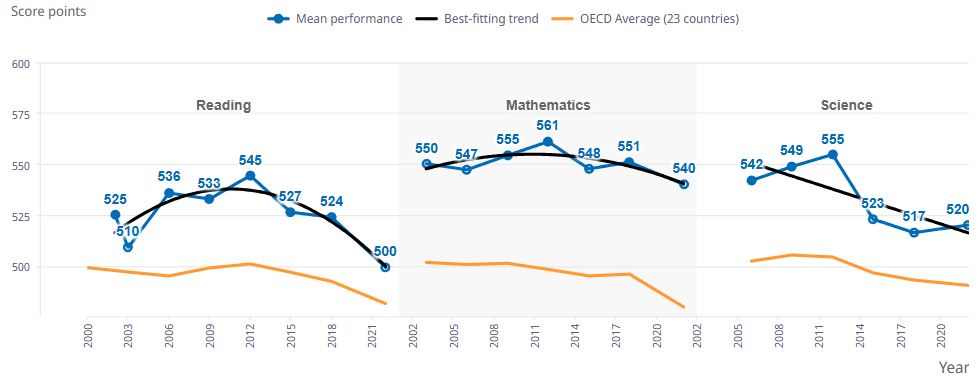


Figure 4. Trends in math, reading and science performance Hong Kong, China.

In the most recent period (2018-2022), the gap between the highest performing students (the top 10% with the highest scores) and the lowest performing students (the bottom 10% with the lowest scores) widened in math and science, while it did not change significantly in reading. In mathematics, the decline in performance was concentrated among low-achieving students, while the results of high-achieving students remained virtually unchanged.

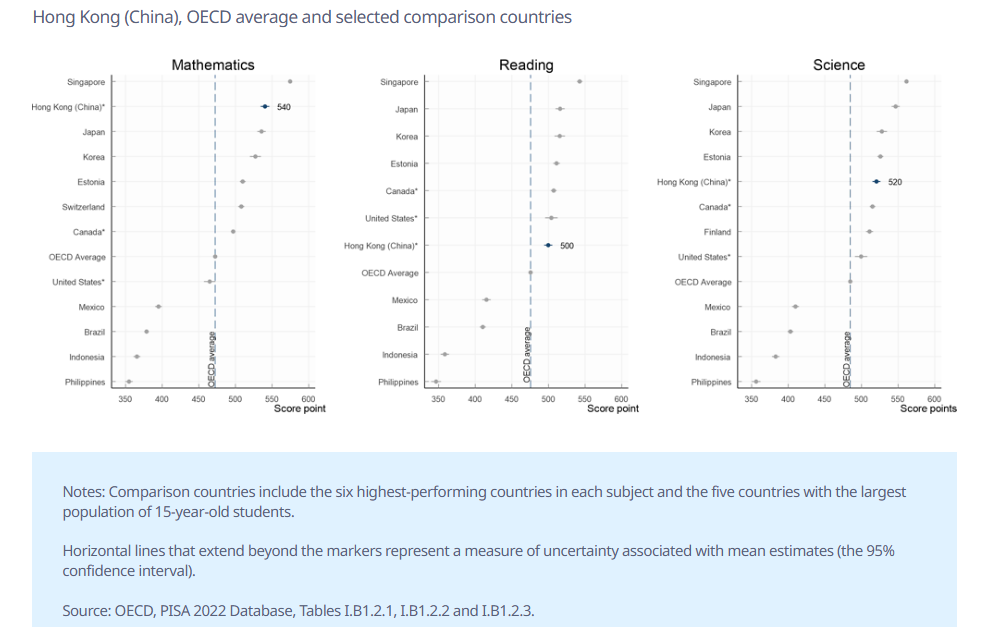


Figure 5. Average results in math, reading and science in PISA 2022

Compared to 2012, the proportion of students performing below the basic level (level 2) increased by five percentage points in math, 11 percentage points in reading, and seven percentage points in science.

Hong Kong students performed above the OECD average in math, reading and science. A higher proportion of students in Hong Kong achieved the highest levels of proficiency (Level 5 or 6) in at least one subject than the OECD average. At the same time, a greater proportion of students achieved the minimum proficiency level (level 2 or above) in all three subjects than the OECD average.

Compared to 2012, the proportion of students performing below the basic level (level 2) increased by five percentage points in math, 11 percentage points in reading, and seven percentage points in science.

In Hong Kong, 86% of students achieved at least level 2 in math, well above the OECD average of 69%. These students can interpret and recognize, without direct instruction, how to mathematically represent a simple situation, such as comparing the total distance of two alternative routes or converting prices to another currency.

Approximately 27% of students achieved the highest scores in the PISA math test, scoring at Level 5 or 6, well above the OECD average of 9%. Six Asian countries and economies had the highest proportion of students at these levels: Singapore (41%), Chinese Taipei (32%), Macau (China) (29%), Hong Kong (China) (27%), Japan (23%) and Korea (23%). Students at these levels are able to model complex situations mathematically, and to select, compare, and evaluate effective strategies to solve problems. In only 16 of the 81 countries and economies that participated in PISA 2022, more than 10% of students reached Level 5 or 6.

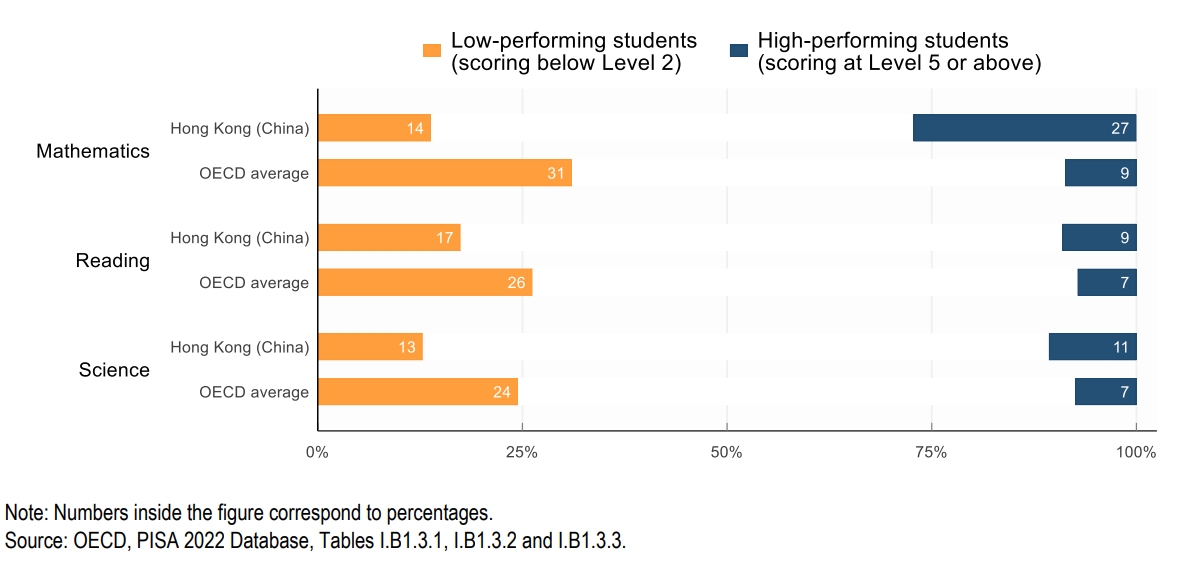


Figure 6. Top and bottom performers in math, reading and science

Approximately 83% of students in Ukraine achieved level 2 or higher in reading, well above the OECD average of 74%. These students can identify the main idea in medium-length texts, locate information based on clearly defined criteria, even if they are sometimes complex, and analyze the purpose and form of texts on request. The proportion of 15-year-old students who achieved minimum reading levels (level 2 or above) ranged from 89% in Singapore to 8% in Cambodia.

In Hong Kong, 9% of students achieved level 5 or above in reading, higher than the OECD average of 7%. These students are able to comprehend long texts, deal with abstract or counterintuitive concepts, and distinguish facts from opinions based on implicit cues about the content or source of information.

Approximately 87% of students achieved level 2 or higher in science, well above the OECD average of 76%. These students can recognize correct explanations for familiar scientific phenomena and use their knowledge to determine the validity of conclusions in simple cases.

In Hong Kong, 11% of students achieved the highest level of science performance, corresponding to Level 5 or 6, compared to the OECD average of 7%. These students are able to apply their science knowledge creatively and independently in a variety of situations, including unfamiliar ones.

These results show that Chinese students continue to achieve high standards in education, which is a great achievement for the country.

While China's PISA results are impressive, they also spark debate. One of the biggest criticisms is that the results represent only the four richest regions of China: Beijing, Shanghai, Jiangsu, and Zhejiang. This is a testament to the high quality of education in these regions, but it means that the results may not reflect the overall picture of education in the country.

PISA results are used by the Chinese government and educational institutions to analyze the strengths and weaknesses of the education system. This contributes to the continuous improvement of educational policy and the implementation of educational reforms. China is actively using PISA data to develop new methodological approaches to teaching, including more practical and research-based teaching methods.

China has participated in PIRLS since 2001, represented by regions such as Hong Kong and Macau. The results of students from Hong Kong and Macau are often impressive in their consistency and high performance. Compared to other countries, Chinese students are usually among the top performers. They demonstrate the ability not only to understand what they read, but also to analyze and evaluate textual information. This is evidence of the effectiveness of the educational system and its ability to adapt to the demands of the times.

In 2016, China showed impressive results, ranking 1st among the participating countries. Students from China scored an average of 590 points, one of the highest scores among all countries. This was a huge achievement for the Chinese education system.

With each PIRLS cycle, China continues to improve its reading achievement by utilizing the latest methods and technologies.

Participation in PIRLS also contributes to teacher professional development. By using research data, educational institutions can better prepare teachers to teach reading skills, taking into account international best practices

It also allows the country to obtain valuable information about the level of students' reading abilities, which helps to improve the quality of education and develop effective educational strategies. Using the results of the study, China can identify problem areas and improve its educational programs.

China has participated in TIMSS since 1995, represented by regions such as Hong Kong and Macau. Here are some key aspects of this participation: Students from Hong Kong and Macau typically perform well on TIMSS tests. They demonstrate the ability not only to solve problems but also to analyze and evaluate scientific concepts. TIMSS results are used to analyze the strengths and weaknesses of students' learning. For example, China analyzes in which aspects its students lag behind their peers from other countries to develop appropriate educational programs and policies. TIMSS data helps to develop strategies to improve learning. For example, Hong Kong has implemented programs aimed at improving math and science skills through the integration of modern technologies and teaching methods

**CHAPTER 4. Prospects for the development of educational measurement in China, challenges and problems**

### 4.1. Prospects for the development of educational measurement in China: innovations in educational measurement

Innovations in educational measurement are an important aspect of the development of the educational system in China. With rapid changes in technology, economy, and society, the need to adapt assessment methods to new realities is becoming increasingly important.

Special attention should be paid to the use of artificial intelligence (AI). AI has the potential to revolutionize educational measurement. In the 2020s, China began to widely use artificial intelligence (AI) in educational measurement. The idea is to personalize each student's learning through adaptive technologies. AI is able to analyze individual student achievements, identify their weaknesses, and offer specialized training programs to improve results. For example, the **Squirrel AI** platform is one of the best examples of AI application in education. It is used for adaptive testing and provides individualized recommendations for each student, taking into account their progress and level of knowledge. Teachers can also receive detailed analytics on the performance of their students, which allows them to customize the learning process more precisely. Another AI-based platform**, iFlytek**, uses technology to assess students' language skills. It helps learners improve their language skills by automatically correcting their pronunciation, grammar, and vocabulary using audio analytics.

Interactive assessments that incorporate gamification elements can increase student engagement and make the assessment process more fun. Using games and simulations for assessments can help students learn better.

Adaptive assessment systems that change the difficulty of tasks depending on the student's answers allow for a more accurate assessment of the level of knowledge and skills.

The introduction of virtual and augmented reality technologies in educational measurements opens up new opportunities for assessing practical skills. For example, students can be exposed to simulations of real-life situations, allowing them to assess their skills in a close-to-real-life environment. These technologies can also be used to create interactive tests that allow students to interact with learning material in a new way.

Although the GCSE is an integral part of China's educational system, there are plans to reform the exam to reduce stress among students and make it more inclusive. As part of the Gaokao reform, attention is being paid to the psychological state of students and their ability to cope with stress. The introduction of psychological tests to help identify stress and excessive pressure on students is an important part of Gaokao's adaptation to modern requirements. The first steps in this direction have already been taken in some Chinese schools, where students are offered psychological counseling before important exams. In the future, it is planned to integrate psychological assessment into the testing process itself.

Using data analytics to assess student performance is becoming increasingly popular. Collecting and analyzing data on student performance allows you to identify trends, problems, and opportunities for improving the learning process. This can include analyzing test scores, classroom participation, and other metrics, allowing teachers and administrators to make informed decisions about instructional strategies.

The growing globalization of education facilitates the exchange of experience and best practices between countries. China can innovate in educational measurements by studying international standards and adapting them to its own conditions.

Cooperation with international organizations and participation in global educational initiatives can help China introduce new assessment methods that are in line with global trends.

This includes participation in international programs such as PISA, TIMMS, PIRLS, and TALIS (Teaching and Learning International Survey), which helps China assess its achievements internationally. Implementation of joint educational projects with other countries, which allows for the adaptation of best practices to the Chinese education system. Existing exchange programs for teachers and students also contribute to the improvement of teaching methods and knowledge assessment.

Prospects for the development of educational measurement in China look promising due to the active introduction of new technologies, international cooperation, and continuous improvement of assessment methods. These factors will help improve the quality of education and ensure equal access to it for all students.

**4.2. Challenges and problems in educational measurement in China**

China's education measurement system faces numerous challenges that need to be addressed and resolved. Insufficient adaptation to new technologies, inequalities in access to education, data security issues, and psychological pressure on students are important aspects that can affect the effectiveness of educational measurement. To succeed in this area, it is necessary to develop strategies that take these challenges into account and contribute to a more equitable and effective educational process.

Let's look at the main problems.

* **Insufficient adaptation to new technologies**:

***Technical infrastructure:*** While cities such as Beijing and Shanghai have developed technological infrastructure, many rural areas of China have serious problems with access to the Internet and modern technology. This limits the ability to implement online testing and automated assessment systems.

***Teacher training***: Many teachers are not sufficiently trained to use new technologies in the educational process. This can lead to uneven implementation of innovations in different schools and regions, as well as to the lack of effectiveness of new assessment methods.

***Resistance to change****:* Some educational institutions may be conservative in their approaches and not ready to embrace new technologies. This may be due to traditional teaching and assessment methods that have been used for many years.

* **Inequality of access to education**:

***Socio-economic factors***: There is significant inequality between urban and rural areas in China, which affects access to quality education and technology. Students in rural areas often have limited access to modern learning materials and technology, which can negatively impact their performance.

***Financial barriers****:* The cost of technology, such as computers and the Internet, may be out of reach for many families, especially in rural areas. This creates additional barriers for students who want to take advantage of new teaching and assessment methods.

***Differences in the quality of education***: There are significant differences in the quality of education in different regions of China. Schools in wealthy urban areas often have better resources, more qualified teachers, and modern technology, while schools in poorer regions may lack funding and resources.

* **Data security issues**:

The growing use of technology in education also raises the issue of data security. Collecting and processing students' personal information requires high standards of privacy and security. Improper data management can lead to data leaks and misuse.

* **Psychological pressure on students**:

The growing emphasis on standardized tests and technological assessment methods can lead to increased stress and pressure on students. This can have a negative impact on their mental health and overall well-being.

### 4.3. Recommendations for improving the system of educational measurement

Based on the analysis of current trends and international experience, several predictions and recommendations can be made for the further development of educational measurement in China:

***Improve technical infrastructure****:* Invest in the development of technological infrastructure, especially in rural and remote areas, to provide access to the Internet and modern technology. This could include expanding broadband access and providing schools with the necessary equipment.

***Teacher training***: Develop professional development programs for teachers that emphasize the use of new technologies in teaching and assessment. This could include trainings, workshops, and online courses to help teachers learn new methods and tools.

***Adaptation of educational programs***: Implement adaptive curricula that take into account the individual needs of students. This may include the use of technology to create personalized learning pathways and assessments that allow students to learn at their own pace. Implement systems that allow for an individualized approach to learning for each student, taking into account their strengths and weaknesses.

***Ensure equal access to education***: Develop programs that promote equal access to quality education for all students, regardless of their socioeconomic status or place of residence. This may include scholarships, financial aid, and support programs for students from low-income families.

***Introduce new assessment methods***: Develop and implement new assessment methods that emphasize critical thinking, creativity, and practical skills. This may include projects, group work, and interactive assignments that allow students to demonstrate their knowledge in real-world situations.

***Ensuring data security***: Develop clear policies and procedures to protect students' personal data. This includes implementing encryption technologies, regular security audits, and training staff on data protection.

***Parent and community involvement***: Actively engage parents and the community in the assessment and learning process. This may include organizing meetings, workshops, and information campaigns to help parents understand the importance of educational assessments and their role in supporting their children.

***Monitoring and evaluation of the system***: Regularly monitor and evaluate the effectiveness of the education measurement system. This may include analyzing test scores, surveying students and teachers, and studying international experience to identify opportunities for improvement.

Improving China's educational measurement system will require a comprehensive approach that takes into account technological, social, and educational aspects. Implementation of these recommendations can contribute to a more effective, accessible, and inclusive assessment system that meets the needs of a modern society.

**CONCLUSIONS**

Given global trends in education and technological advances, China is actively developing its educational measurement system. Given the need to improve the quality of education, adapt to international standards, and strive for innovation, the country is trying to preserve the strengths of its educational system while responding to the challenges faced by students, teachers, and society.

The historical development of educational measurement in China reflects a complex interplay between tradition and innovation. From imperial examinations to modern standardized tests and international research, the system has changed in response to political, economic and social demands. The implementation of educational reforms at different stages of history has allowed China to achieve high results in education and ensure its high quality.

In many developed countries, objective information obtained in national and international research is used to make political decisions.

Data on student performance obtained in the course of national and international assessments of the quality of education are used to monitor the performance of different subgroups of the study population and help to allocate resources more evenly.

The results of such studies can spark a broad political discussion and initiate changes in the education system.

Education assessment programs themselves (and not just their results) are often used to motivate change by informing the public about the state of the education system, familiarizing teachers with different assessment systems and opportunities to adjust curricula.

The prospects for the development of educational measurement in China are encouraging, in particular due to the active implementation of technology, inclusive approaches and reforms in the Gaokao system. China is striving to adapt its educational measurement to modern requirements, while maintaining high efficiency and responding to global trends. It is expected that in the coming years, China's educational measurement system will become more flexible, inclusive and technologically advanced, which will contribute to the development of not only academic achievement, but also creative and social skills of students.

**LIST OF REFERENCES**

1. Dai Xinlan. Exploration of the status quo and realisation path of China's education modernisation [N]. Jiangsu Economic News,2024-08-30(T04).

2. Zhou Changhu. Milestones in the History of China's Educational Development [J]. Current Affairs (《Current Affairs Report》Middle School Students Edition), 1994, (01):25.

3. Liu Changya. Accelerating the modernisation of education and starting a new journey to build a strong education country - Interpretation of China's Education Modernisation 2035 [J]. Education Research, 2019, 40(11):4-16.

4. Chen Xing,Zhang Xuemin. Educational input in new China: Criteria, methods and indicators of evaluation and their transmutation [J]. Tsinghua University Education Research,2019,40(02):89-98.DOI:10.14138/j.1001-4519.2019.02.008910.

5. Sun Xiao Jian, Song Nai Qing, Xin Tao. Implementation and Implications of Multi-Stage Adaptive Testing in PISA Testing. Tsinghua University Press. 2021. No 5. P. 72-78.

6. Wei Ya Qin, Wang Jin Wen, Luo Ma. Introduction to the TIMSS-2023 Mathematics Assessment Framework and its Implications for Mathematics Education Evaluation in Compulsory Education in China, Monthly Teaching Magazine (Secondary School Edition). 2023. No 1. P. 3-8.

7. PISA. Programme for International Student Assessment. PISA - PISA (oecd.org)

8. Jiang Tianhui, Li Zihua. International Comparison of PISA. Transmission and the Role of Data Governance from a Neoliberal Governability Perspective. Tsinghua University Educational Research. 2023. P. 103-110.9. Chunlian Jiang, Do-Hong Kim,Chuang Wang. Analysis of a Mathematical Problem-Solving Test on Speed and Students' Strategies: A Mixed Item Response Theory Approach // Chinese/English Journal of Educational Measurement and Evaluation. 2021. Vol. 2: Iss. 3, Article 3.

10. Liang Zhengyan; Zhang Minqiang; Huang Feifei; Kang Derong; and Xu Lingling. Application Innovation of Educational Measurement Theory, Method, and Technology in China’s New College Entrance Examination Reform // Chinese/English Journal of Educational Measurement and Evaluation.2021. Vol. 2: Iss. 1, Article 3.

11. Zhang Houcan and Luo Fang. The Development of Psychological and Educational Measurement in China // Chinese/English Journal of Educational Measurement and Evaluation. 2020: Vol. 1: Iss. 1, Article 7.

12. Lijia Guo, Jiashun Huang and You Zhang. Education Development in China: Education Return, Quality, and Equity // Sustainability 2019.

13. Liao, C., & Livingston, S. A search for alternatives to common-item equating. // Paper presented at the annual meeting of National Council on Measurement in Education, Vancourver, British Columbia, April, 2012.

14. Wei, Y., Liu, J., & Dorans, N. J. Evaluation and exploration of linking design for a language test // Unpublished manuscript. 2013

15. Wei, Y., & Morgan, R. An evaluation of the single-group growth model as an alternative to common-item equating (Research Rep. 16-01). 2016. Princeton, NJ: ETS.

16. The Definition and Selection of Key Competencies: Executive Summary, OECD, Paris, http://www.oecd.org/pisa/35070367.pdf (accessed on 05 April 2018).

17. JIANG Enlai; LI Yong; QIN Hongxia. From Leadership to Overall Leadership of the Communist Party of China:The Triple Logic of the Development of Higher Education under the Leadership of the Communist Party of China [Elektroniс resource]: kns.cnki.net

18. XIE Saiyin. The Influence of Educational Modernization on the Teaching Reform of Ideological and Political Courses in Colleges and Universities [Elektroniс resource]: kns.cnki.net

19. Communist Party of China:The Triple Logic of the Development of Higher Education under the Leadership of the Communist Party of China [Elektroniс resource]: kns.cnki.net

20. ZHOU Hongyu & LI Yuyang On the Construction of High-quality Education System [Elektroniс resource]: kns.cnki.net.

21. Tierney, R. D. Fairness in educational assessment. In M. A. Peters (Ed.), Encyclopedia of Educational Philosophy and Theory. 2016. Singapore: Springer Science+Business Media.

22. Jun Li, Jian Li. Educational Policy Development in China in the 21st Century: A Multi-Flows Approach // beijing international review of education 1 (2019) 196-220.

23. Snapshot A. Education in China. OESD, 2016. 64 р.

24. <http://timss.bc.edu>

25. <http://www.pisa.oecd.org>

26. https://pirls.bc.edu/

27. <https://nus.org.ua/wp-content/uploads/2020/02/PISA2018_Mizhnarodnyj-zvit_ukr.pdf>

[28. https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-i\_5f07c754-en](28.%20https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-i_5f07c754-en)

29. <https://reading.cw.com.tw/trend-article/83>

[30. https://ouci.dntb.gov.ua/en/works/4zenGKv7/](30.%20%20https://ouci.dntb.gov.ua/en/works/4zenGKv7/)

31.<https://www.researchgate.net/publication/237441809_China_Challenges_of_Secondary_Education>

32.<https://www.researchgate.net/publication/298911898_Standardized_Testing_In_The_Case_Of_China_And_The_Lessons_To_Be_Learned_For_The_US>

33.<https://www.researchgate.net/publication/248939170_Educational_Assessment_in_China_Lessons_from_history_and_future_prospects>

34. <https://journals.openedition.org/ries/3846?lang=en>

35. <https://www.tandfonline.com/doi/abs/10.1080/09695940120033216>

36. <https://repository.library.georgetown.edu/handle/10822/924637?show=full>

37. <https://ouci.dntb.gov.ua/en/works/7pnJ3vBl/>

38. <https://nus.org.ua/articles/pisa-2022-korotkyj-oglyad-usih-osnovnyh-rezultativ/>

39. PISA 2022 RESULTS (VOLUME I) - OECD 2023