Chapter 9
Assessing Science Learning in Schools: Current Policy and Practices

Cuidian Feng and Lingbiao Gao

9.1 Context of School Science Assessment

China is known as the "kingdom of examinations." Various tests and exams are set for students during their school time. With the trend of "exam-oriented education" getting more powerful, only pen-and-paper tests were used in all kinds and all levels of exams in science after the early 1990s. These tests and exams used to be achievement-oriented, elitist, and bureaucratic (Gao 1998). The power of examination also made it "a baton conducting teachers, students, and the teaching-learning process" which led to an exam-oriented style of teaching and learning in all schools throughout China. Teaching and learning focused sharply on drilling students with exam techniques in order to get higher marks (Gao and Watkins 2002). In this way, assessment became an obstacle for improving the quality of learning and teaching. It stood on the opposite side of research findings in education and worldwide trends of science education and curriculum reforms in the past decades (Gao 2002).

To keep pace with the world, China moved to reform its national curriculum for basic education, as well as the school assessment system, and changed its orientation toward personal quality development. According to the *Guidelines for Curriculum Reform in Basic Education* (*Trial*) (hereafter referred to as *Guidelines* in this chapter) published by the Chinese Ministry of Education (MOE) in 2001, the aims of this reform pursue to change the curriculum:

C. Feng (✉)
School of Teachers Education, Taizhou University, Linhai, Zhejiang 317000, China
e-mail: fengcuidian@163.com

L. Gao
South China Normal University, 55 Zhongshan Dadao, Tianhe, Guangzhou 510631, China

L.L. Liang et al. (eds.), *Chinese Science Education in the 21st Century: Policy, Practice, and Research*, Contemporary Trends and Issues in Science Education 45, DOI 10.1007/978-94-017-9864-8_9

- From a knowledge-delivery center to fostering students' all-around development, including knowledge, processing skills and methods, emotions, attitudes, and values
- From a subject-based and decollated structure to a well-balanced, comprehensive, and properly flexible structure that able to meet the needs of all students
- From course contents that were overloaded and overly difficult to ones that have appropriate amounts of material and degree of difficulty and which can enhance the relations between course contents and students' lives as well as keep pace with progress in science and technology
- From too much emphasis on rote learning to higher emphasis on meaningful learning
- From testing for ranking and selecting students to assessing mainly for facilitating students' learning and development
- From an overly centralized and unified system to a more diverse and flexible system which allows schools and local governments to share the responsibility in curriculum development and management (MOE 2001a)

New policies and techniques of assessment were adopted in order to:

> build up a new assessment system aimed at facilitating students' all-around development. It will not only assess students' learning achievement, but also discover and develop students' potential in a variety of ways, identify their needs in progress, and help them to develop their self-understanding and self-confidence. Assessment needs to play its roles in educating students and facilitating their development. (MOE 2001a, Guideline No.14)

A new wave of assessment reform has spread over China since the release of the above *Guidelines* in 2001. In 2002, the Ministry of Education released another document named *Circular of the Ministry of Education on Promoting Reforms on School Evaluation and Examination System* (hereafter referred to as *Circular*) to elaborate further the aims and philosophy of the assessment reform:

(a) The reform should follow the education policy of the Chinese Communist Party, which aims at assessing and facilitating students' all-around development morally, intellectually, physically, and aesthetically.
(b) School evaluation system, including both student assessment and teacher evaluation, is not only for administration purposes, but more importantly, for the development of students, teachers, and schools. The most important purpose of this reform is to change traditional school evaluation into a facilitative process for the development of students, teachers, and schools in terms of quality of learning, teaching, and educating.
(c) The contents of assessment should be diverse in order to cover all aspects of student development rather than subject knowledge only. Great attention needs to be paid to keeping a good balance between the uniform curriculum standards and personal difference in students' capacity and personality.
(d) Improvements on assessment techniques and instruments are encouraged in order that pen-and-paper tests are not the only tools used in assessment. Summative and formative tests, quantitative and qualitative techniques, and

intrinsic and extrinsic assessments all should be included in the new system. Intrinsic assessments and self-assessments are especially encouraged. Students and teachers are no longer treated only as objects to be evaluated but also as subjects to assess their own progress.

(e) Attention should be paid not only to the results of assessment but also to changes and progress that have occurred in the process of assessment.

(f) The importance of the roles of students, teachers, and schools in the process of evaluation should be noted. It is expected that assessment will become a process of interaction among students, teachers, schools, education officers, and administration sections as well as parents (MOE 2002).

The *Circular* also drafted a framework of the objectives of the new student assessment system. Assessment objectives were classified into two groups: (1) "general development objectives" in six major domains of student development as shown in Table 9.1 and (2) "subject-based learning objectives" in three dimensions of learning outcomes, which were described in a series of national curriculum standards for all school subjects.

The *Circular* also mandated a portfolio named "records of progress" that to be established for every student to promote ongoing assessment focusing on the process of learning. The portfolio is supposed to collect qualitative information, i.e., the students' self-records of their learning and schooling, peer evaluation results, the students' best work, students' performance in community service, awards received by students in any competition, teachers' observations and comments, comments from parents, etc. It should give an all-around description and deep understanding of the students' development in the process of learning. Furthermore, it will promote the students' reflection about their learning process and encourage students to play a more active role in self-assessment.

At the end of each term and school year, a qualitative assessment focused on the "general literacy and capacity" of students, as shown in Table 9.1, is conducted. Reports including students' achievement test scores at the end of the term or year, their self-reports on learning, peer evaluations, and teachers' comments are put together and sent to students' parents to give an all-around assessment on the progress of students. Final reports of a student's "general literacy and capacity" at the end of junior or senior secondary stages also play a role in the acceptance of students into senior high schools or universities. In the junior secondary stage, general scores are given to students based on the results of assessment of their "general literacy and capacity." This score is 10 % of the final score for senior secondary school entrance. Another 90 % of the score comes from the results of public examinations. In the senior secondary stage, information provided by the report on "general literacy and capacity" becomes one of the basic requirements for college acceptance. Tests and examinations are still the major approach for summative assessments at midterm (in the senior secondary stage, the end of a module), at the end of term, and at the end of school year. A rating scale such as "excellent, good, or satisfactory" or "pass or fail" is recommended to grade students' achievement instead of the popular percentage marking scale in primary schools. Ranking

Table 9.1 The general development objectives of students in six major domains

Domains	Indicators
Ethics and morality	Love our motherland, our people, and our socialist system. Be fond of work. Respect for laws. Be honest. Be public spirited and concerned with the community and environment
Civic literacy	Have self-confidence, self-esteem, self-reliance, self-discipline, and diligence. Be responsible for personal behavior and the society. Be active in activities for public welfare
Learning ability	Have aspiration and interest in learning. Be able to learn in a variety of ways and learn at higher level. Be accustomed to thinking reflectively on the process and outcome of one's own learning. Be able to analyze and solve problems by applying the knowledge, skills, and experience learned in different areas. Develop abilities of inquiry and creativity
Capacity of communication and cooperation	Be able to set and complete a task with others. Be respectful and receptive to the views and conditions of others. Be able to evaluate and regulate one's own behavior. Be cooperative and able to communicate and interact with others in a variety of ways
Participation in physical activities and conditions of health	Love to participate in sports. Regularly take part in physical exercises. Practice and be skillful in sports. Keep the body strong and healthy and live in a healthy way
Awareness of beauty	Be able to enjoy and appreciate the beauty of life, nature, art, music, and science. Have a good and positive aesthetic taste. Be active in a variety of art and music activities, and be able to perform in variety of ways

Based on the Circular (MOE 2002)

students according to their scores on the tests is not suggested in primary school stage. The primary graduation examination and the junior secondary school entrance examination are eliminated. Primary graduates are distributed randomly to nearby junior secondary schools in their communities. At the end of the 9-year compulsory education, the junior secondary graduation examination and the senior secondary entrance examination are merged into one regional public examination. The university matriculation examinations are decentralized and become public examinations at the provincial level, though they are still named as the national university matriculation examinations.

In sum, the MOE *Circular* drew a blueprint of the expected assessment system, which aims at facilitating the all-around development of students (and so was named "developmental assessment system"). This new system is characterized by (a) a diversity of its acting personnel and (b) a diversity of assessing methods and techniques. It seems to agree with the concepts of "assessment for learning" (Black and Wiliam 1998) and "assessment as learning" (Eral 2003) that have acknowledged increased attention and importance internationally. In the past decades, it has been argued that assessment has to move from "assessment of learning" to

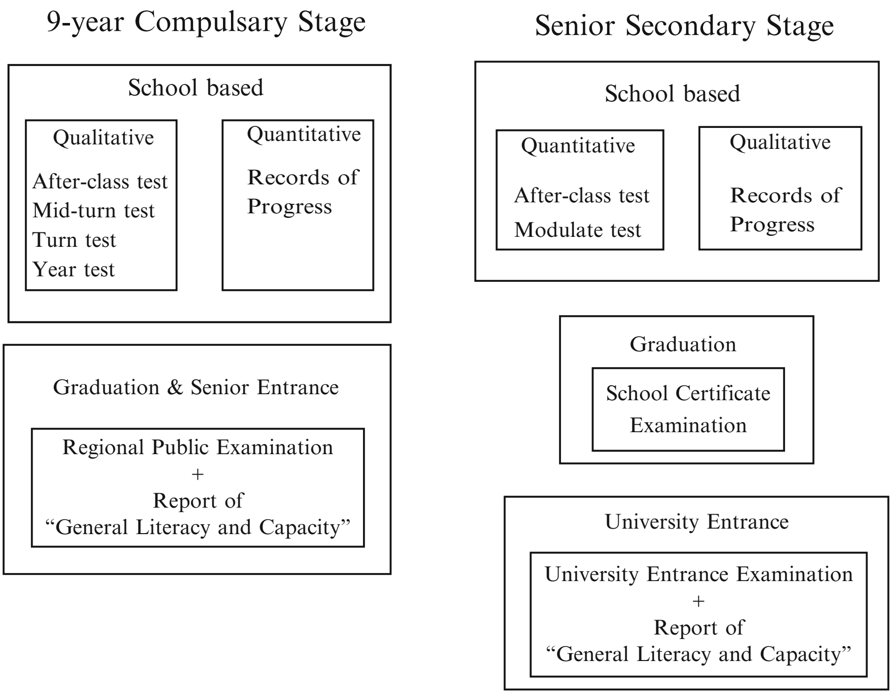

Fig. 9.1 School assessment system in China

"assessment for learning," where assessment procedures and practices are developed to support learning rather than undermine learning (Gipps 1999; Shepard 2000; Stiggins 2002). In addition, student-involved assessment, i.e., assessment as learning, has received more emphasis and recognition as a core component where assessment can support learning, (Eral 2006; Noonan and Duncan 2005; Torrance 2007). Figure 9.1 gives a brief summary of the new school assessment system in China today (Gao 2013).

9.2 Distinctive Features of the Current Assessment Reform

As mentioned above, China has reconstructed its national curriculum and released a reform to change its assessment system accordingly. In order to have a deep understanding of the new system of learning assessment, it is necessary to understand how this reform is distinct from the former assessment system. These distinctions also served as indicators for the review and inspection of the success of the reform.

Firstly, the aims of assessing student learning are different. Traditionally, student assessment in China is aimed at inspecting and evaluating student learning in terms of students' performance on summative pen-and-paper tests. The new policy advocates that student assessment will inform student learning and development by identifying their progress and diagnosing their problems in the process of learning. Not only are the progress and problems in cognitive field learning assessed but also those related to student learning and development in all-around aspects.

Secondly, the focus on the contents of assessment is different. Traditional assessments focus only on knowledge, especially textbook knowledge. Under the new policy, objectives of assessment cover three dimensions, as suggested in the *Circular*: (a) knowledge, (b) processing skills and methods, and (c) emotion, attitudes, and value (MOE 2002). This suggests that not only will scientific knowledge be assessed, but students' understanding of scientific inquiry; inquiry skills and methods; capacity for applying scientific knowledge, skills, and methods in scientific inquiry and everyday problem solving; and understanding of the underlying philosophy of scientific inquiry will all be included in learning assessment. In addition, students' emotions about, attitudes toward, and value of science are also parts of assessment content. This enlarges the scope and strengthens the role of learning assessment. Gao (2004) recognized this as the highlight of the assessment reform and the most significant distinction between the new and the traditional assessment systems.

Details of the contents of learning assessment are further defined in the national curriculum standards developed by MOE. They are subject dependent and varied with school stages. At the primary level, the main areas of assessment in science include (MOE 2001b):

(a) Children's knowledge and understanding of the basic scientific concepts in the living world, the physical world, the earth, and the space, as defined in the curriculum standards
(b) Children's skills and abilities in "doing" scientific activities
(c) Children's awareness and understanding of scientific inquiry and learning science
(d) Children's interest in and attitudes toward science and learning science
(e) Children's affection for science and the enterprise of science

At the junior secondary stage, the subject courses, physics, chemistry, biology, and geography, are separate in the entire country except for Zhejiang province, which has an integrated science course. At the senior secondary stage, all subject courses are separate all over the country. The details of knowledge contents of learning assessment are different from subject to subject. However, the focus areas of learning assessment are similar for all subjects and at both the junior and senior stages. According to the junior secondary school curriculum standards of science, physics, chemistry, biology, and geography (MOE 2001c, d, e, f, g) and the senior secondary school curriculum standards of physics, chemistry, biology, and geography (MOE 2003a, b, c, d), learning assessment focuses of on the following areas:

(a) Students' understanding of the scientific inquiry process and its key steps: challenging and questioning, assuming, planning, experimenting and collecting evidence, analyzing and concluding, reporting and communicating, examining, and reviewing; the capacity for completing learning activities relevant to one or several of the above steps
(b) Students' understanding of the scientific knowledge defined by the curriculum standards, especially the big ideas that underlie the scientific concepts
(c) Students' skills in learning and doing science, such as skills in observing, measuring, experimenting, operating instruments and tools, and data collecting and analyzing
(d) Students' ability to apply scientific knowledge and solve problems
(e) Students' interests in and attitudes to science and learning
(f) Students' understanding about the nature of science and scientific enterprise
(g) Students' understanding and views on STS issues

Thirdly, with the changes in the focus areas in assessment, the strategies of assessment need to change accordingly. The new policy encourages a variety of means of assessment rather than relying on pen-and-paper tests only. At the primary school stage, it strongly emphasizes that strategies of learning assessment in science should (MOE 2001b):

(a) Develop carefully to meet the aims of the new curriculum
(b) Focus sharply on the indicators defined by the curriculum standards
(c) Handle the level of difficulty properly to meet the level of student learning
(d) Develop carefully and apply properly the new techniques and instruments for assessment

It should be noted that a diversity of qualitative techniques are introduced in the primary science curriculum standards. These include classroom observation, interviews, after-class learning tasks, DIY (do-it-yourself) activities, project learning activities, student portfolios of progress, and performance tests. Self-assessment and peer assessment of students are encouraged. It is suggested that teachers should pay attention to their reflective language and behaviors while communicating or interacting with students since these are specific ways of assessment that happen subconsciously. Furthermore, summative pen-and-paper tests are not compulsory for all students; local educational authorities can decide independently whether or not to conduct summative tests at the end of a period of learning. This means that quantitative and summative examination is no longer the major approach to learning assessment in primary schools. Thus, qualitative and process-oriented assessment techniques become the mainstream of learning assessment strategies in primary school science.

At secondary school stage, there is little difference in assessment strategies adopted by different subjects and at different periods. According to the national secondary school science curriculum standards documents (MOE 2001a, b, c, d, e, f, g, h), the strategies of learning assessment in secondary science will:

(a) Serve to diagnose and improve student learning.

(b) Involve multiple personnel in the assessment. Not only teachers, external experts, and administrators but also the parents and students could all act as assessors.
(c) Limit the scope of knowledge in assessment and control of the level of difficulty properly.
(d) Develop and apply a variety of flexible techniques and instruments in learning assessment.

It can be seen from the national curriculum standards that summative pen-and-paper tests are still emphasized in the secondary school stage. However, all the curriculum standards call for a change in designing test items to shift the focus from memorizing knowledge to understanding the scientific concepts and the underlying big scientific ideas. The use of open-ended items is encouraged in summative tests, although they might cause difficulty in marking. In addition, the importance of internal assessment is recognized due to the overemphasis of external assessments in past decades. This is different from the policy at the primary stage where summative pen-and-paper tests are not encouraged.

The use of performance tests is encouraged in the secondary school stage. Operational skills in observation and experiments, including skills in handling tools and instruments, identifying and focusing on objects, collecting data and evidence, and properly recording the observed information and data measured, have become the major content of performance tests. Students' self-assessment and peer assessment are also encouraged in performance tests.

The record of progress is introduced as one of the most important instruments in learning assessment for improving students' learning autonomy. It acts as not only an instrument of assessment but also a platform for interaction and communication among students, teachers, and parents and a good way for students to review their own learning. The strategy applied here is similar to that in the primary stage. However, the techniques employed in different stages must be different to fit the level of maturity and development of students.

Qualitative techniques, such as observations and interviews, as already described at primary stage, are also introduced in order to make a change in assessment.

9.3 Internal Science Assessment in Practice

As described in Fig. 9.1, internal assessment in science includes processing assessments in and after class and the summative assessments after a certain period of learning, i.e., after a module, half-term, a term, or a school year. Primary graduation tests are still in practice in most primary schools in China. According to the *Circular*, these tests ought to be school-based. However, in reality, most of the primary graduation tests are run by district/county educational authorities to compare the quality of schools (Gao 2013), so these tests are only for school accountability purposes rather than for student accountability. Thus, such graduation tests and examinations are excluded here and will be addressed in Sect. 9.4.

9.3.1 Portfolio Assessment in Practice

Portfolio assessment was first introduced by the Ministry of Education in the *Circular* (MOE 2002). Every student in China has to work with a "record of progress" when schooling. It was considered a form of authentic assessment in China. Authentic assessments offer multiple indicators of student progress and encourage students to take an active role in their own learning and to demonstrate what they know in ways that encompass their personal learning styles. It increases students' ownership of course content, provides a first step in researching, and offers more opportunities for writing, discussion, and the use of technology. Independent learning and creative problem solving are also encouraged. A move toward more authentic application tasks and outcomes thus improves learning and teaching, and students have greater clarity about their obligations. Teachers can thus come to believe that assessment results are both meaningful and useful for improving instruction (Wiggins 1993).

As a popular instrument, the portfolio requires students and teachers to document students' growth and changes by selecting evidence from their teaching and learning practices. This will help students to become more self-regulated and gain personal control and independence in their learning. They will be able to use a wide variety of learning styles to demonstrate their learning. They will also be able to develop a greater understanding of their particular learning style when they self-evaluate and reflect on the evidence they have selected for inclusion in the portfolio to demonstrate competence.

Neither the central government nor the local educational authorities give a clear definition of the construction and contents of this "record of progress." As a result, a diversity of "portfolios" have been developed by teachers and educators in the past decade. A review of the portfolio used in practice shows that there are three main kinds of portfolios (Zhao 2012):

1. *Demonstrative portfolios*. These are self-selected collections of students' best work chosen in order to demonstrate their achievements to teachers, parents, and fellow students. Students also explain the criteria and reasons for the selections. Teachers, parents, and other students give their comments on both the works demonstrated and the criteria for selection. This type of portfolio could work as a medium for student-teacher-parent communication. It could also be a good instrument to help students to learn how to assess themselves by setting assessment criteria and selecting their best work.
2. *Descriptive portfolios*. This is a systematic collection of a variety of records in the process of students' learning. It includes the teachers' evaluations, observations, and comments on students' performance in the learning process, the

students' achievement in tests or examinations, and the students' own works and any other records the students or their teachers regard as significant to collect. Both the students and their teachers have the right to add things into the portfolio. This portfolio is very important for students (or their teachers) to review their learning processes, identify their progress, and diagnose their problems.

3. *Summative portfolios*. This is a collection of a student's work selected by teachers and educators to report the achievement of that student to parents and the society. Preset standards are necessary and need standardization to maintain fairness for all students. The content of this type of progress record normally covers the six major domains of student development as shown in Table 9.1.

Since most of the portfolios are not subject-based, they facilitate students' all-around development, including their learning in science. In the National Forum on Assessment of the General Literacy of High School Students run in Xian in October 2010, teachers from schools in different provinces summed up the advantages of portfolio assessment:

1. Records of progress are a very rich resource of information about students. They give detailed descriptions of the all-around aspects of students in their process of maturity and development.
2. Records of progress can act as media promoting interactions between teachers, students, and their parents and can therefore benefit students' development.
3. Records of progress help teachers to learn the individual differences among students and the features of each student. Teachers can then give each student more appropriate guidance according to the characteristics shown in the portfolio.
4. Records of progress provide platforms for students to submit the work that they themselves are satisfied with and fond of. This enables students to see their own progress and experience the joy of success and progress and, in turn, improves students' attitudes toward and emotion about learning.
5. Records of progress highlight the active role of students in the process of assessment. They provide chances for students to review their own learning.
6. Records of progress collect a vast amount of qualitative and quantitative evidence of student development for summative assessment. This provides an effective way to integrate summative and formative assessments and to integrate teaching, learning, and assessment.

Many teachers enjoy their experience in portfolio assessment process and believe that the records of progress facilitate student progress in the right direction. Mr. Li Weifeng, a high school physics teacher at the Dongzhimen Middle School in Beijing, said[1]:

[1] Extracted and translated from Li Weifeng: *A speech on the National Forum on Assessment of the General Literacy of High School Students*. Unpublished manuscripts of the forum, Xian, China, October 2010.

. . . Student portfolios record the details of students' progress in the past three years. . . . I was very impressed and excited as I read their portfolios page by page. I saw that my students grew up day after day. They are happy to learn and to enquire. They are enthusiastic and hardworking. They are highly concerned with our society. They are kind, honest, wise, and intelligent. They are the future of China. . . .

. . . I am very surprised by and proud of my students when I review their progress. I am sure that they would not be able to progress so well if we focused only on their knowledge learning and achievement and if we are concerned only about their performance on the public examinations.

What makes them progress so well? ...... Records-of-progress do not focus narrowly on subject learning and exam marks. Rather, the teachers' comments, the comments from their peers, the experience of thinking reflectively, and self-evaluation: all of these lead them to progress actively in the right direction.

Similar to teachers, many students enjoy their experience in portfolio assessment process and view the portfolio as a record of life. A student in the First Middle School of Yinchuan, the capital city of Ningxia Autonomous Area in Western China, wrote down her feeling on the record of progress:

When I reviewed my experience in the past three years, I found this small thing [the portfolio] had changed me. I became more self-confident and more focused on learning. Furthermore, I saw a new me. . . .It is a condensation of my life. Every detail in the portfolio is like a drop of water reflecting my life and it is so beautiful. I am very impressed by my own portfolio. I am so proud and happy with my experience.[2]

Some of the parents were also happy with their kids' experience, which is reflected actively in the portfolio. A parent wrote to her daughter in the portfolio:

Half a semester has passed since you entered senior high school. We are very happy to see that you are progressing with joy. Meanwhile, we are impressed with your collection [in the portfolio]. Remember, what you have collected is not only a resume; further, it is the record of your progress and growth. It will enrich your life. Your learning is excellent in that you learn actively with self-consciousness. We hope that you will keep on learning in this way. This will help you progress well. Go ahead. Your future is bright.[3]

However, this is only one side of the picture. On the other side, problems remain unsolved in portfolio assessment. Firstly, conducting portfolio assessment is a heavy workload for teachers. This is partially due to the big class size in China. A typical class in primary schools has 40–50 students and in secondary schools 50–60. If a teacher spends 10 min per week to read and give feedback to one student, 500–600 min needs to be spent on portfolio assessment per week. This is a very heavy workload for the teacher. Many teachers are enthusiastic in the beginning but then begin to feel tired of doing portfolio assessment. Secondly, due to the very strong impact of public examinations, quite a large number of teachers still consider exam marks to be the most important result of schooling. They view

[2] Extracted and translated from the portfolio of Miss LYX, a portfolio shown at the National Forum on Assessment of the General Literacy of High School Students, Xian, China, October 2010.

[3] Extracted and translated from Zhao Xueqin: *A speech at the National Forum on Assessment of the General Literacy of High School Students*. Unpublished manuscripts of the forum, Xian, China, October 2010.

portfolios as less significant and not worth spending a lot of energy on. Similar problems trouble the students also. Many students, especially those with lower motivation to learn, feel that it is too tiresome to do portfolio assessment. They feel bored in collecting their own works and are not happy to show them to others, especially to parents. How to encourage teachers and students to be actively involved in portfolio assessment becomes a big problem.

These problems trouble educators and teachers and stand in the way of implementing portfolio assessment in schools. In fact, only a small number of teachers and schools, mainly in the primary stage, are involved actively in using records of progress. In most schools, especially at the secondary stage, portfolio assessment becomes a mere formality.

9.3.2 Performance Assessment in Practice

One popular type of performance assessment in China refers to a practical learning task, such as a science experiment, an inquiry learning activity, or a real or analogous problem-solving task. Through observing, recording, and analyzing the examinee's behaviors, teachers can assess not only operational skills, cognitive level, and thinking skills but also the desire and capacity for learning, cooperation, and communication. Students can also assess themselves and their peers. The history of performance assessment in China is rather long; however, due to the very strong influence of high-stakes public examinations which use only pen-and-paper tests, performance assessment has been neglected for many years. The *Circular* reemphasizes the importance and value of performance assessment, so it has become popular again, especially in primary and junior secondary stages.

Summarized from the Chinese literature in recent years, a typical performance test includes three key stages:

(a) *Preparation*. This stage focuses on setting a task with its learning objectives and criteria for assessment. The task and the corresponding objectives and criteria might be assigned by the curriculum or a textbook. In other cases, the task is assigned by the teacher, and the teacher sets the objectives and criteria for learning and assessing beforehand.
(b) *Assessment*. This stage includes the student's self-assessment, peer assessment, and teacher's assessment. The students' self-assessment and peer assessment focus on reviewing and rethinking their approaches and behaviors to find out what they achieved, the mistakes they made and the problems they met in the process, and their feelings and ideas about the activity, learning, and science. The teacher's assessment focuses on students' performance by observation and communication. As an example, Xu and Yan (2008) describe the details of assessing students' performance in an inquiry activity in chemistry. In order to help students to conduct self-assessment, the teacher negotiated with them and then finally constructed a rubric consisting of the criteria shown in Table 9.2.

Table 9.2 A rubric for student self-assessment in an inquiry activity

Title of the task:		
Name:	Gender:	Class:
Points to review		Your description
What is your general feeling about this experiment?		
Did you have your own assumptions before you started to conduct the experiment? If yes, did you make a plan to improve your assumption?		
Did you have problems in the process of experimenting? What are the problems? How did you solve the problems?		
How did you work as a team in the experiment?		
Did you find you needed to do anything to improve your cooperation?		
Did you stop during the process of experimenting? Why did you stop?		
Did you know how to use the instruments in the experiment? Did you operate them properly?		
Did you ask questions to the teacher when you were not sure about something during the experiment? What were the questions?		
Can you give a comment on this activity? Give reasons to support your comment		
If you need to rank your performance as excellent, good, fair, pass, or fail, what is your rank?		

Translated from Xu and Yan (2008)

Students were invited to write down their feelings and experience with the help of the rubric. Here, a qualitative approach was adopted.

Another rubric was designed for student's peer assessment, which is shown in Table 9.3.

Teachers are still the most important people in assessment since they can assess students' performance more objectively than outside people can. They prepare the criteria of assessment and share it with all students involved in the activity. A rubric of this assessment has been shown in Table 9.4. Students' capacity and skillfulness in designing and conducting an experiment are the focuses of this rubric. It evaluates process skills and results of the activity, identifies the student's capacity, and diagnoses problems the student has in the activity.

(c) *Marking and reporting*. Sometimes, it is necessary to rank or to mark the students' performance. Normally, all the results of students' self-assessments, peer assessments, and teacher assessments will contribute to the final result. However, a descriptive report is always more important than the rank/mark. The indicators shown in Table 9.4 also give an outline describing students' performance.

Most of the teachers accepted the ideas of performance assessment. However, due to the lack of tools and technical assistance, many of them do not know how to conduct performance assessment effectively in practice. This encouraged the

Table 9.3 A rubric for student peer assessment in an inquiry activity

Title of the task:		
Name:	Gender:	Class:
Points to review		Your comment
Was the student actively involved in searching for information relating to the activity?		
Did the student make his/her own assumption before the activity?		
What was the information the student collected? Where did he/she collect that information?		
What kind of work did the student complete?		
Did the student contribute any ideas or suggestions?		
Did the student help other group members when they had problems? If yes, what were the problems?		
Did the student cooperate actively with other group members to fulfill any task? If yes, what was the task?		
Did the student listen to others?		
Can the student express his/her ideas neat and clear while communicating with others?		
What is the result/conclusion of that activity?		
Please give your rank to this activity from one of the five following: excellent, good, fair, pass, and fail		

Translated from Xu and Yan (2008)

Chinese educators and teachers to develop a variety of techniques for performance assessment in recent years, including worksheets and rubrics for science inquiry activities, rubrics for observations, peer assessments and self-assessments, and team competitions.

Table 9.5 presents an example of a marking scale used to assess primary third grade pupils' skills in handling a thermometer (Ling 2009). In this activity, students are required to heat the cold water in a beaker and use a thermometer to measure the temperature of the water while heating. The group leaders are appointed by the teacher and are marked by the teacher beforehand. They also learn marking skills and then act as assessors to mark other members of the group. The group leaders rotate so that each member will have an opportunity to act as a group leader.

A worksheet can also be used as a tool to assess students' performance in this activity. Many researchers encourage the application of worksheets and have developed a large number of worksheets for different activities (Luo 2006; Zhao and Pan 2010; Cai 2012). Table 9.6 gives an example of a worksheet for an experiment on the relations among three variables: electric current, voltage, and resistance (Luo 2006). Students conduct the experiment following the guide in the worksheet step by step, writing down their answers to the questions presented on the worksheet. They report on the process and results of their performance in the activity, and then the teacher can assess the student on their knowledge and experimentation skills based on their written responses and the results of their experiment.

Table 9.4 A rubric for teacher assessment in an inquiry activity

Title of the task:					
Name:	Gender:	Class:			
	Qualitative description			Grade	
Indicators	Excellent	Good	Satisfactory	Pass	Fail
1. Was the student able to put forward proper questions as the focus of the experiment?					
2. Was the student able to put forward hypotheses for the experiment?					
3. Was the student able to provide a reasonable rationale to support the hypothesis?					
4. Was the plan for the experiment reasonably designed?					
5. Were the instruments for the experiment selected properly?					
6. Were the students able to set up the instruments quickly and correctly?					
7. Were the rules properly followed in the operation of the balance in weighing chemical reagents?					
8. Did the student take care to save the chemical reagents?					
9. Was the experiment table clean?					
10. Were the instruments clean?					
11. Were the results of the experiment correct?					
12. Was the way of analyzing reasonable?					
13. How was the attitude of the student in cooperating with others?					
14. How well did the student express her/himself orally?					
15. How actively was the student engaged in the experiment?					
16. Your overall evaluation of the activity					

Translated from Xu and Yan (2008)

Some assessment rubrics refer to scientific inquiry activities (project learning) generally. Jian et al. (2005) gave an example of this kind of rubric as shown in Table 9.7.

After 10 years of practice, many researchers and teachers agree that performance assessment can:

(a) Collect more and deeper information about students' learning
(b) Help understand students' science literacy in all-around aspects
(c) Vary with the level of students and their learning environment
(d) Improve the validity of student assessment
(e) Make science experiments and activities more attractive
(f) Promote students' active and creative involvement in learning

Table 9.5 A scale for student peer assessment of skills in using a thermometer. Grade 3, class, group, assessor (group leader)

Elements	[a]GM1	GM2	GM3	GM4
1. Grasp the upper end of thermometer and put it into the water. The lower end of the thermometer should not touch the bottom or the wall of the beaker				
2. Keep your eyes at the same level as the liquid surface inside the thermometer. The thermometer should not leave the water while reading the temperature				
3. Read and record the temperature of hot water correctly				
4. Read and record the temperature of cold water correctly				

Translated from Ling (2009)
[a]*GM* Group member

Table 9.6 Sample worksheet for a physics experiment

Exploring the relations among electric current, voltage, and resistance
I. The following materials are on your table: three fixed-value resistors, one variable resistor, one ammeter, one voltmeter, two batteries, one switch, and several pieces of wire. Please check these items. Please raise your hand if there is anything missing
II. The question for you to explore, what are the relations among the following variables: electric current, power, voltage, and resistance?
III. Please perform the following tasks:
1. Make a hypothesis about the relations among electric current, voltage, and resistance based on your knowledge and experience
2. Design an experiment to test your hypothesis (complete the electric circuit on the picture of the components. Explain your procedure
3. Conduct the experiment based on your plan. Record the data in the table you have preset
4. Analyze the data and draw your conclusions from the data

Translated from Luo (2006)

(g) Facilitate students' development of their inquiry abilities (Wei 2007)

Similar to portfolio assessment, performance assessment also has shortcomings. The major problems are as follows:

(a) It is time consuming to conduct performance assessment. A good way to solve this problem is to invite students to play the role of assessor. When the assessment focuses on some simple objectives, it is easier for students to handle the criterion of assessment. However, it is not easy for students to handle complex tasks.
(b) Most of the rubrics in the literature are too complicated. The criteria for assessment are not well defined. This increases the difficulty of implementing performance assessments. In addition, the rubrics developed by different people are different even when they focus on the same issue. On one hand, this enables the assessment to fit closely to a certain group of students and their level of

Table 9.7 A rubric for assessing scientific enquiry activities

Elements	Indicators	Marking scale			
		D (fail)	C (accept)	B (good)	A (excellent)
Observing and questioning	Student is able to ask question based on context	Unable to ask question	Ask vague question	Ask surface question	Ask valid scientific question
Predicting and hypothesizing	Student is able to make prediction and hypothesis	Unable to make prediction	Predict the results but incorrectly	Predict the results nearly correctly	Predict the results correctly
Planning and designing experiment	Student is able to plan and design experiment accordingly	Unable to design the experiment	Design a part of the experiment	Design the experiment but not feasible	Design a feasible experiment
Experimenting and collecting data	Student is able to implement the experiment and collect the necessary data	Unable to conduct the experiment	Conduct the experiment with errors and mistakes	Operate correctly but cannot record the data	Operate correctly and record the data exactly
Analyzing and concluding	Student is able to analyze the data and draw conclusions	Unable to analyze the data	Analyze a part of data without conclusions	Analyze data correctly and draw superficial conclusions	Analyze the data and draw in-depth conclusions
Arguing and evaluating results	Student is able to make argument and evaluate the process and results	Cannot evaluate the results	Compare results with the expectation without further consideration	Compare results with the expectation and discover/ raise new questions	Compare results with the expectation, discover/raise new questions based on comparison, and attempt to answer the new question to some extent
Discussing and applying	Student is able to	Achieve none of the 1, 2, 3, 4, 5	Achieve one of the 1, 2, 3, 4, 5	Achieve two of the 1, 2, 3, 4, 5	Achieve three or more than three of the 1, 2, 3, 4, 5
	1. Discuss actively with his/her own opinion				
	2. Write report with convincingness				

(continued)

Table 9.7 (continued)

Elements	Indicators	Marking scale			
		D (fail)	C (accept)	B (good)	A (excellent)
	3. Question others' report scientifically				
	4. Cooperate well with others				
	5. Work creatively				

Translated from Jian et al. (2005)

learning, but on the other hand, different groups who using different assessments cannot be compared. For this reason, applying performance assessments is limited to larger-scale public examinations.

9.3.3 Summative Assessment in Practice

The situation of summative assessment in primary schools is different from that in secondary schools. In the primary stage, since the summative test in science is not compulsory, the situations vary from place to place. For example, in Guangdong province, no summative test or another type of assessment is set for primary school science. Teachers must rank their students at the end of each term and school year. However, since there are only two ranks, pass or fail, set for reporting students' achievement in their term/annual report, almost all students can get a pass except those absent from the science classes. This means that summative assessment does not in fact exist in primary schools in Guangdong. Zhejiang province is an example a different situation. In Zhejiang, most city bureaus give a pen-and-paper test at the end of each term. Since the pen-and-paper test can only test some of the surface knowledge of science at the primary stage, the final score of a student will consist of three parts: (a) students' records of their pen-and-paper test at the end of the school term and year, (b) students' capacity for processing skills and their understanding of scientific methods, and (c) students' emotions about and attitudes toward learning and inquiry. The last two parts are based on the formative/process assessment. Table 9.8 shows a sample marking scheme in a primary school (Ling 2009).

In both the junior and senior secondary stages, formal summative tests are administered at the end of each term and year. End-of-module tests and midterm tests are also administered in every school. Pen-and-paper tests are still the most popular method of summative assessment. Changes have mostly occurred in the techniques for developing test items. Increasing numbers of open-ended items testing students' abilities of comprehension and application are being developed and used in the pen-and-paper tests. Multiple-choice items that only can test trivial

Table 9.8 A summative marking scheme in primary science

Dimensions	Student code:		Class:	
	Grade	Comments	Assessment	Score
Science concepts			End-of-unit test	
			End-of-semester test	
Process and methods			Experimentation skills	
			Checklist of science learning	
			Best student work sample	
Emotion, attitudes, and values			Teacher observation	
			Peer and self-assessment	
			Monthly group-wide competition	
Final score				

Note:
1. Grading scale: a score between 90 and 100 will be given an A (excellent)
A score between 75 and 89 will be given a B (good)
A score between 60 and 74 will be given a C (acceptable)
A score below 60 will be given a D (needs improvement)
2. The weight of the two scores in scientific concepts is 50:50; the weight of the three scores in process and methods is 20:40:40; the weight of the three scores in emotion, attitude, and value is 40:30:30; and the weight of the three-dimensional scores in the final score is 40:30:30

knowledge are declining in tests and exams. In some schools, performance tests are also included in the summative tests. The tasks of performance tests could be (a) an experiment, (b) an academic paper, (c) a DIY product, (d) an oral presentation or demonstration, etc. (Gao 2011). A final score considering both the results of summative tests and process assessment is given to students to show their achievement in learning. Table 9.9 gives an example of the composition of a final score in senior physics (Zhang 2005).

From Table 9.9, one can see that summative assessment includes both pen-and-paper tests and performance tests. There are two kinds of pen-and-paper tests: one is the traditional closed-book test and the other is an open-book test. In the open-book test, students may be asked to write an academic paper on an issue relating to the knowledge learned, draw a concept map related to the concepts learned, present an argument on a STS issue, etc. Most of the performance tests ask students to conduct an experiment they have done in the term. Most of the unit tests in the process assessment use a pen-and-paper test to assess students' knowledge. Qualitative techniques are used in other parts of process assessment. These include students' self and peer assessments and teacher's assessment. It should be noted that attendance is also a factor in the assessment. Students must attend the course and accumulate enough class hours in order to get credit. Otherwise, they cannot get credit in that course even if they get a very high mark on the final test.

It should also be noted that not all the secondary schools assess student learning in the way shown in Table 9.7. In fact, many schools continue to use closed-book pen-and-paper tests as the only method of assessing students' learning in science, which is nearly the same as the situation before the new curriculum reform.

Table 9.9 A term report of student achievement in physics

	Elements and contents			Results of assessment				
Process assessment	Unit tests		Unit 1	Mark		Rank		Comments
			Unit 2	Mark		Rank		Comments
			Unit 3	Mark		Rank		Comments
			Unit 4	Mark		Rank		Comments
	Attitude and method		Learning attitude		Excellent	Good	Acceptable	Needs to improve
			Learning habit					
			Learning method					
	Cooperation and communication		Respect for others		Excellent	Good	Acceptable	Needs to improve
			Easy to get along with others					
			Express correctly					
	Peer assessment							
					Signature of group leader			Date
	Self-assessment							
						Signature of student		Date
	Teacher assessment							
			Final rank (pass/fail)			Signature of teacher		Date
Summative assessment	Pen-and-paper tests	Close	Mark		Rank		Comments	
		Open	Mark		Rank		Comments	
	Performance test		Mark		Rank		Comments	
Total # of hours required for attendance						Total # of hours actually attended		
				Signature of teacher				Date

9.4 External Examinations in Science

As shown in Fig. 9.1, no official public examination is mandated in science at the primary stage. At the end of the junior secondary stage, a public examination is required as the graduation exam as well as the senior secondary school entrance exam. This exam is now a criterion-referenced exam named the junior secondary school certificate examination (JSSCE). There are two public examinations for senior secondary students: the high school certificate examination (HSCE), held in grades 11 and 12, and the university matriculation examinations (UMEs). The UME is the most important and high-stakes examination in China.

9.4.1 Features and Issues Related to Junior and Senior Secondary School Certificate Examinations (JSSCEs and HSCEs)

Science is one of the subjects of the JSSCE. The exam paper for science consists of three parts, physics, chemistry, and biology, which relate to the three separate science subjects in junior secondary schools. According to the Ministry of Education, JSSCE should use a variety of methods to assess students' levels and capacities comprehensively and include the results of processing assessment. A five-point or seven-point rating scale has been introduced as a substitution for the traditional percentage scale (MOE 2008). Since the JSSCE is organized by city or regional bureau of education, the overall situation in the country is too complicated to give a detailed description. However, a brief review on the JSSCE web page[4] shows that almost all senior high schools admit students solely based on their exam marks. In science, pen-and-paper tests are still the most important method of the JSSCE,[5] and the percentage score is still the only scale for marking students' performances. This suggests that in the past decade, there has been no significant change in external science exams in the junior secondary stage. The design of exam items, however, now seems to be more ability-oriented (Liao and Yuan 2010).

At senior secondary stage, the HSCE is run at the provincial level. Wu (2012) reviewed the HSCE in 19 provinces and found that science is one of the testing subjects in all the HSCE. In some provinces, science appears as a single testing subject, and in some provinces, it appears as three separate subjects. In some provinces, performance tests that focus on experimentation skills are now included in the exams (see Table 9.10). However, she found that the HSCEs are immature and underdeveloped in terms of their purposes, schemes, objectives, methods, and interpretations. She identified two major problems in the HSCE:

1. The nature of the HSCE is unclear. It should be a criterion-referenced test, but the existing form is a mixture of criterion- and norm-referenced examinations. The role of the HSCE is to confirm that the student meets the curriculum standards upon graduation. However, the current HSCE focuses partly on selection and overlaps with the national university matriculation examinations.

4 http://gz.zhongkao.com/

5 A few cities and provinces include listening and oral exams in the examinations for foreign languages. A few cities and provinces adopt open-book exams as a method in exam of political science.

Table 9.10 Methods used by different provinces in different subjects of HSCE

Subject	Student[a]	Method[b]	Provinces
Physics, chemistry, biology	All	Ext. test	5 provinces
	S & L		2 provinces
	All	Ext. test + experiment. 1	6 provinces
	S & L		1 province
	All	Ext. test + exp. 2	13 provinces
	S & L		1 province
	All	Ext. test + exp. 3	1 province

Notes:
Exp. 1 = external test of experiment, run by the educational authority
Exp. 2 = internal test of experiment, run by the school
Exp. 3 = internal test of experiment, run by the school + daily performance
[a]All = all students, S & L = students majoring in social science or liberal arts
[b]Ext. test = external pen-and-paper test

2. The criteria/standards of the HSCE tests are missing in the schemes of almost all provinces or are at least not open to the public.

These two serious problems might cause the HSCE to become invalid, unreliable, and unnecessary. Wu (2012) suggests that it is very urgent that the Chinese education authorities, at both national and provincial levels, clearly define the nature and roles of the HSCE in order to develop and publish the criteria/standards for the HSCE.

9.4.2 Features and Issues Related to University Matriculation Examinations (UMEs)

There have been a few changes in the UME. First, the UME has been decentralized. Only one UME scheme, developed by the National Education Examination Authority (NEEA), was used in the country before the national curriculum reform. The test was also developed solely by the NEEA and was the same everywhere in China. This was unfair, since the levels of social and economic development and the quality of education are different from one region to another. With the implementation of the new national curriculum, the administration of the UME has been moved down to the provincial level, and UME exam schemes are proposed by provincial educational examination authorities; the tests are developed by a group of university experts appointed by provincial education examination authorities.

Second, test items have been improved. Traditional pen-and-paper tests can assess only students' knowledge memory and lower-level thinking skills. A large number of experts and teachers have criticized this. However, the pen-and-paper test is still the only method used for the science portion of the UME. This has caused more people involved in these studies to focus on improving the quality of pen-and-paper test items in order to test students' concepts understanding and higher-level thinking skills. Currently, a variety of new items simulating the real scientific research situations and problems have been created. Students are requested to design a scheme for an experiment for inquiry or to solve a set questions. A deep understanding of knowledge, higher thinking, and problem-solving skills are needed to answer these items. This makes the UME highly challenging to most students.

There are other small changes too. However, these improvements have not significantly changed the basis of the UME.

9.5 Summary and Discussion

School science assessment in China has been characterized as examination-dominated, achievement-oriented, and selective in nature. It has been criticized as an obstacle for the cultivation of students' scientific literacy. When the MOE launched the innovation to develop a curriculum that is more student-centered, ability-focused, and aims at developing the all-around person for schools in China, assessment reform became urgent. It started soon after the national curriculum reform at the beginning of the new century. However, there are big gaps between the policy and practice.

On the policy level, a review of the government documents shows that the MOE has tried to build a new assessment system that aims at facilitating the all-around development of students. In this new system, the acting personnel are diverse: not only teachers and external experts can act as assessors, but also students can do so. The focuses of learning assessment extend from scientific knowledge to the all-around development of scientific literacy in three dimensions: scientific knowledge and skills, ability of scientific inquiry and creativity, and emotions about and attitudes toward science and the value of science. Self-assessment and peer assessment are encouraged. A variety of ideas, methods, and techniques have been introduced. The orientation of these changes agrees with the concepts of "assessment for learning" and "assessment as learning."

On the practical level, the ideas of process assessment, the introduction of the records of progress, and other qualitative techniques have become the most distinctive changes in science learning assessment, especially in primary stage when the wash-back effects of the high-stakes public examinations are not so strong. In addition, a variety of performance test techniques have been developed to be used in both formative and summative assessment. These new ideas, methods, and techniques have extended the scope of students, teachers, and, especially, parents on students' all-around development. They also combine students' learning and assessment into one process. Student self-evaluation and peer evaluation techniques are being widely used in the process of learning, which encourage students to review and learn from their own experiences of learning.

In secondary schools, especially in senior secondary schools, formative assessment and records of progress have not been as popular as expected. Summative tests are still the most prevalent mode of student assessment. The introduction and application of performance tests seem to be the most important change. This might be due to the fact that secondary school students have to face three public examinations, the JSSCE, HSCE, and UME. Only a few changes have occurred in the past decade in those public examinations, and those changes have concentrated on improving the quality of test items. Different kinds of new test items have been developed and tried in order to test not only students' knowledge memory and surface understanding but also their high-order thinking skills and problem-solving abilities. These have made the public examinations increasingly difficult and have also seemed to push students to work harder and focus more specifically only on the academic aspects of development. This is the opposite of curriculum reform's desired effect of cultivating students in all-around aspects. It also creates powerful wash-back effects which pull the students and teachers back to the traditional way of learning.

Why are the practical situations of assessment reform different from the original expectation? Gao (2013) reviewed the challenges in student assessment reform in the past decade:

> Firstly, the policy is top-down and the interpretations of this policy from different levels are confusing. For example, different experts and administration agencies at different levels interpret the concept "developmental assessment system" differently. Secondly, the qualitative assessment techniques are young and not well developed. More importantly, public examinations in China are still one of the most important ways to achieve social fairness and flexibility. All these strengthen the traditional routine of assessment. So in practice, the implementation of the assessment reform has not gone as smoothly as expected. Significant problems exist, which has made the reform progress slowly, and to some extent, drift off its original direction and become the so-called "bottleneck" of China's education innovation. (p. 455)

Student assessment in school science learning has undergone a series of changes toward a more student-centered, learning-facilitated, and development-oriented in China. However, assessment of learning is still the most dominant approach in schools since it fits the needs of social fairness and flexibility. There is still a long way to make the dreams of "assessment for learning" and "assessment as learning" come true in science learning assessment. It depends not only on the efforts of educators and teachers but also on changes in Chinese society.

References

Black, P. J., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College.

Cai, W. Y. (2012). The classroom assessment based on worksheet in process of science inquiry instruction (科学探究教学中基于工作单的课堂评价). *Contemporary Education Science, 2*, 25–28 (In Chinese).

Eral, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks: Cowin.
Eral, L. M. (2006). *Rethinking classroom assessment with purpose in mind. Manitoba Education, Citizenship, and Youth*. Canada: MECY.
Gao, L. B. (1998). Cultural context of school science teaching and learning in China. *Science Education, 82*(1), 1–14.
Gao, L. B. (2002). Issues in student assessment in the new national curriculum (新课程的评价问题). *Global Education, 31*(6), 52–56 (In Chinese).
Gao, L. B. (2004). The concept of processing evaluation and its function (过程性评价的功能和理念). *Journal of South China Normal University (Social Science Edition), 6*, 102–106 (In Chinese).
Gao, L. B. (2011). *Rethinking student assessment reform theoretically* (学生评价改革的理论思考). A lecturer on the national workshop for senior science teachers, 12/12/2011. Guangzhou: South China Normal University. (In Chinese).
Gao, L. B. (2013). Reforms in student assessment in Mainland China. In E. H. F. Law & C. Z. LI (Eds.), *Curriculum innovations in changing societies: Chinese perspectives from Hong Kong, Taiwan and Mainland China* (pp. 449–472). Rotterdam: Sense Publishers.
Gao, L. B., & Watkins, D. A. (2002). Conceptions of teaching held by school science teachers in P.R. China: Identification and cross-cultural comparisons. *International Journal of Science Education, 24*(1), 61–79.
Gipps, C. (1999). Social-cultural aspects of assessment. *Review of Research in Education, 24*, 355–392.
Jian, Y. G., Xian, H. Z., & Jin, P. (2005). The construction and practice of the assessment system of scientific inquiry teaching (科学探究教学评价体系的构建与实践). *Curriculum, Teaching Material and Method, Beijing China, 25*(12), 60–64 (In Chinese).
Liao, B. Q., & Yuan, L. M. (2010). The analysis of characteristics of science inquiry ability assessment in physics Senior-high School Entrance Examination in 2009 (2009年中考物理试题对探究能力考查的特点探析). *Basic Education Curriculum, 3*, 71–74 (In Chinese).
Ling, B. (2009). Ideas and practice of primary science learning assessment (小学科学课教学评价的思考与实践). *Science Class (Primary Edition), 10*, 20–21 (In Chinese).
Luo, G. Z. (2006). Study on performance assessment of scientific inquiry abilities (对科学探究能力表现性评价的研究). *Curriculum, Teaching Material and Methods, 26*(8), 66–69 (In Chinese).
Luo, G. Z. (2007a). Research of gender difference of performance evaluation in scientific exploration (科学探究表现性评价的性别差异研究). *Theory and Practice of Education, 27*(5), 41–42 (In Chinese).
Luo, G. Z. (2007b). Study on effectiveness of assessment methods of science inquiry ability based on paper-and-pencil test (基于纸笔测验的科学探究能力评价的有效性研究). *Shanghai Education Research, 10*, 51–54 (In Chinese).
Luo, G. Z. (2007c). The assessment methods of science inquiry abilities (科学探究能力的评价方法). *Education Science, 23*(3), 7–10 (In Chinese).
Ministry of Education (MOE). (2001a). *Guidelines for curriculum reform in basic Education (trial)* (基础教育课程改革纲要(试行)). (Official document, No.17) (In Chinese).
Ministry of Education (MOE). (2001b). *Primary science curriculum standards for compulsory education (3–6) (trial) (科学课程标准(3–6年级)(实验稿))*. Beijing: Beijing Normal University Press (In Chinese).
Ministry of Education (MOE). (2001c). *Junior secondary science curriculum standards for compulsory education (7–9) (trial)* (科学课程标准(7–9年级)(实验稿)). Beijing: Beijing Normal University Press (In Chinese).
Ministry of Education (MOE). (2001d). *Physics curriculum standards for compulsory education (trial)* (义务教育物理课程标准(实验稿)). Beijing: Beijing Normal University Press (In Chinese).

Ministry of Education (MOE). (2001e). *Chemistry curriculum standards for compulsory education (trial)* (义务教育化学课程标准(实验稿)). Beijing: Beijing Normal University Press (In Chinese).

Ministry of Education (MOE). (2001f). *Biology curriculum standards for compulsory education (trial)* (义务教育生物课程标准(实验稿)). Beijing: Beijing Normal University Press (In Chinese).

Ministry of Education (MOE). (2001g). *Geography curriculum standards for compulsory education (trial) (义务教育地理课程标准(实验稿))*. Beijing: Beijing Normal University Press (In Chinese).

Ministry of Education (MOE). (2001h). *Guidelines for kindergarten education* (*trial*) (幼儿园教育指导纲要(试行)). No. 20, 2001. (In Chinese).

Ministry of Education (MOE). (2002). *Circular of the ministry of education on promoting reforms on school evaluation and examination system* (关于积极推进中小学评价与考试制度改革的通知). (*Official document,* No. 26). (In Chinese).

Ministry of Education (MOE). (2003a). *Physics curriculum standards for senior secondary schools (trial)* (普通高中物理课程标准(实验稿)). Beijing: People's Education Press (In Chinese).

Ministry of Education (MOE). (2003b). *Chemistry curriculum standards for senior secondary schools (trial) (普通高中化学课程标准(实验稿))*. Beijing: People's Education Press (In Chinese).

Ministry of Education (MOE). (2003c). *Biology curriculum standards for senior secondary schools (trial)* (普通高中生物课程标准(实验稿)). Beijing: People's Education Press (In Chinese).

Ministry of Education (MOE). (2003d). *Geography curriculum standards for senior secondary schools (trial)* (普通高中地理课程标准(实验稿)). Beijing: People's Education Press (In Chinese).

Ministry of Education (MOE). (2008). *Suggestions on promoting and improving the reform of senior secondary school entrance examinations* (关于深入推进和进一步完善中考改革的意见). No. 6, 2008. (In Chinese).

Noonan, B., & Duncan, R. (2005). Peer and self-assessment in high school. *Practical Assessment, Research & Evaluation, 10*(17), 1–8.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*, 758–765.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education, 14*(3), 281–294.

Wei, F. S. (2007). Performance assessment in process of primary science instruction (小学科学教学中的表现性评价). *Basic Education Research, 5*, 39–41. Nanning China, In Chinese.

Wiggins, G. (1993): *Assessment, authenticity, context, and validity*. Phi Delta Kappan, November, 200–214.

Wu, Y. (2012). *A review on the reform of high school certificate examination in China*. Paper presented at the 8th conference of International Test Commission, Amsterdam, 3–5 July.

Xu, M. M., & Yan, M. G. (2008). Designing performance assessment for enquiring experiment in chemistry (化学探究性实验活动表现评价的设计探讨). *Journal of Chemical Education, 10*, 12–13 (In Chinese).

Zhang, J. P. (2005). *Instruction and assessment of senior physics* (物理教学与学业评价). Guangzhou: Guangdong Education Press. (In Chinese).

Zhao, B. G. (2005a). The characteristics of the assessment on scientific inquiry ability in GCE Physics in England (英国GCE物理课程科学探究能力评价的特点). *Curriculum, Teaching Material and Methods, 25*(9), 93–96 (In Chinese).

Zhao, B. G. (2005b). Experimental and inquiry ability—implication from GCE Physics test (实验与探究能力的评价—来自GCE物理考试的启示). *Journal of Physics Teaching, 23*(6), 1–2 (In Chinese).

Zhao, D. C. (2012). Review and reflection on the application of portfolio (成长记录袋应用的回顾与反思). *Curriculum, Teaching Material and Method, Beijing China, 32*(5), 21–26 (In Chinese).

Zhao, M. L., & Pan, S. D. (2010). Worksheet: An instrument for valid assessment of scientific process skills (工作单:一种有效评价科学过程技能的工具). *Education Research of Jiangsu, 1A*, 46–48 (In Chinese).