Chapter 10
Alignment Between the National Science Curriculum Standards and Standardized Exams at Secondary School Gateways

Xian Chen, Minyan Jiang, Li Cai, Ling L. Liang, Jing Du, and Yan Zhou

10.1 Introduction

China is well known for its exam-oriented education system, especially at the secondary school level. Although the most recent curriculum reform was launched a decade ago, progress in transforming secondary school science instruction has been slow and unproductive nationwide. In reality, the goal of “quality education” is generally acknowledged, but test preparation often overrides national curriculum standards (Zhao 2009). Many teachers continue to adopt traditional frameworks of science disciplines and teacher-centered instructional approaches in their classrooms. To ensure more successful implementation of standards-based science education, the following two aspects appear to be most critical: (1) reform the national college entrance exam and college admission system and (2) ensure consistency between the existing standardized exams at gateways and the national curriculum standards (Yuan et al. 2002). If certain topics and/or cognitive skills are

---

X. Chen (✉) • L. Cai
College of Teacher Education, Nanjing Normal University, 122 Ninghai Road, Nanjing,
Jiangsu 210097, China
e-mail: cxwzx2006@163.com

M. Jiang
School of Physics and Technology, Nanjing Normal University, 1 Wenyuan Road, Nanjing,
Jiangsu 210023, China

L.L. Liang
La Salle University, Philadelphia, PA, USA

J. Du
Nanjing Foreign Language School, 30 Beijing East Road, Nanjing, Jiangsu 210008, China

Y. Zhou
Jiangyin First Middle School, 88 Changjiang Road, Jiangyin, Jiangsu 214431, 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_10

consistently neglected in the exams at these gateways, the teachers are more likely to ignore those topics or cognitive skills in classroom teaching despite the curriculum standards requirements.

In this chapter, we focus our research on the second aspect mentioned above through examining the degree of alignment between the national science curriculum standards documents and the standardized exit exams at junior and senior high schools. We first briefly review the existing alignment research literature within and outside of China and then report the findings based on our own research studies. At the end of the chapter, we discuss implications and make recommendations for further research.

10.2 Alignment Research and Its Significance

In an effective educational system, important components such as content standards, curriculum, instruction, and assessment should be well aligned to send a consistent message about what is valued in the educational process (Webb 1999). Alignment research presents one way to evaluate the level of agreement or consistency between those abovementioned educational aspects or components. Alignment is also getting more attention as an aspect of the validity and reliability of an assessment and its uses (Beck 2007). Results of an alignment study may help policymakers, assessment developers, curriculum developers, and educators make informed decisions on further refinements of educational systems.

According to Bhola and colleagues (2003), alignment can be defined as the extent of agreement between an academic standards document and the assessment (s) used to measure student learning of these standards. Alignment can be measured by multiple models that vary in levels of complexity. At the simplest level, an alignment study might only identify the content match between the assessment and a specific set of curriculum standards. At highly complex levels, multiple aspects of alignment of standards and assessments would be examined. For instance, one of the most widely used alignment procedures used by individual states within the United States, Webb's model (1997, 1999) analyzes alignment using four criteria: (1) categorical concurrence, indicating the extent to which the test content is consistent with corresponding content standards; (2) depth-of-knowledge consistency, indicating whether the test content matches the specified level of cognitive challenge in the standards; (3) range-of-knowledge correspondence, indicating the extent to which the content in the standards is covered in an assessment; and (4) balance of representation, indicating the degree to which the test's content distribution is balanced across objectives and consistent with the standards. In Webb's alignment model, a criterion for each of the four measures is given. If all four criteria are met, then the alignment between the test and the corresponding standards is claimed to be acceptable (Webb 2007) (Table 10.1).

One may argue that the more dimensions involved in an alignment model, the more the findings drawn from the model would more accurately represent the

Table 10.1 Characteristics of the physics exit exams at the senior high school level by province

	Guangdong	Jiangsu	Shandong	Hainan	Ningxia
Feature	Economically well-developed region -------------------------→ economically underdeveloped region				
Time (min)	90	75	90	60	100
Total points	100	100	100	100	100 +20 (bonus)
Format	Multiple choice questions only	Multiple choice, fill-in-the-blank, word problem involving mathematical calculations or discussion/argumentation involving real-life situation	Multiple choice, fill-in-the-blank, word problem involving mathematical calculations	Multiple choice, fill-in-the-blank, experimentation-related word problem involving mathematical calculations	Multiple choice, fill-in-the-blank, graphing word problem involving mathematical calculations, discussion/argumentation involving real-life situation
Bonus question	No	No	No	No	Yes (20 points)

reality. However, it is also true that the more complex an alignment model is, the more difficult a high alignment between the assessment items and the standards can be achieved (Bhola et al. 2003; Fulmer 2011). As an example of a moderately complex model of alignment, Porter (2002) focuses the alignment analysis on both content and cognitive domain's dimensions using the data collected through the Surveys of Enacted Curriculum (SEC) (Council of Chief State School Officers 2004). Unlike the Webb model, the SEC approach does not rely on direct comparison of assessment items with content standards. Instead, a content taxonomy is first developed as the common framework by subject matter experts, and then trained content analysts map the standards and assessment items onto the taxonomies using two-dimensional matrices—one dimension for content and the other for cognitive demands, which allows differentiation of levels of content difficulties. The categories of Porter's cognitive demands are consistent with the revised Bloom's taxonomy for describing learning outcomes (Anderson and Krathwohl 2001). To evaluate the level of alignment, the matrices for standards and assessments are compared by cell, and an alignment index, P, is calculated to indicate the proportion of content in common (see Tables 10.2 and 10.3). The use of a common language and the quantitative alignment index produced in Porter's model allows the level of alignment to be calculated and compared across different standards documents, assessments, textbooks, classroom instruction, and many other components of the educational system. In addition, based on a simulation study, Fulmer (2011) further established critical values for Porter's alignment index, suitable for hypothesis testing at alpha levels of 0.05 and 0.10.

In a recent study, Polikoff et al. (2011) investigated the coherence of standards-based reform in the United States by analyzing 138 standards-assessment pairs spread across grades and the three tested subjects required by law. With the SEC approach, it was found that roughly half of standards content was covered on the corresponding test and roughly half of test content corresponds to the standards. Misalignment also occurred due to a mismatch on cognitive demands between the assessment items and the standards content. About 17–27 % of content on a typical test covers topics not mentioned in the corresponding standards. The authors again argued that in order to ensure that all students experience equal learning opportunities and demonstrate their achievements, the standards, assessment, and instruction must work together to deliver a consistent message.

10.2.1 Alignment Studies in China

Alignment research in China is relatively new. The first alignment study about standardized exams at high schools in China was a case study published in 2008, based on an analysis of the 2002 Chinese National Physics Syllabus (Grades 10–12) and the alignment between the guidelines and two recent twelfth-grade exit tests in Jiangsu Province (Liang and Yuan 2008). The data were also later used in an international comparison study of the physics curriculum standards and the physics

Table 10.2 Physics content standards (grades 10–12) by topic and by cognitive level

Topic	Sub-topic	Remember	Understand	Apply	Analyze	Evaluate	Create	Subtotal
Describing motion (DM)	Significance of experiment and mathematics method, physics model, particle model, displacement, speed, acceleration, law of uniform rectilinear motion	2(0.03)	2(0.03)	1(0.01)	0(0.00)	0(0.00)	0(0.00)	5(0.07)
Interaction forces and laws (IFL)	Static and kinetic friction forces, Hooke's law, forces in two dimensions, vector and scale, Newton's laws, overweight, weightlessness, the standard SI unit	3(0.04)	4(0.06)	1(0.01)	2(0.03)	0(0.00)	0(0.00)	10(0.14)
Mechanical energy and sources of energy (MESE)	Work, mechanical power, kinetic energy theorem, gravitational potential energy, the conservation of mechanical energy, energy in various forms and conservation	5(0.07)	8(0.11)	0(0.00)	1(0.01)	0(0.00)	0(0.00)	14(0.19)
Projectile and circular motion (PCM)	Projectile motion, circular motion, centripetal acceleration and force	1(0.01)	1(0.01)	1(0.01)	3(0.04)	0(0.00)	0(0.00)	6(0.08)
Classical mechanics achievement and limitation (CMAL)	Law of universal gravitation, artificial satellite, the second and third universe speed, Newton's and Einstein's views about gravitation, scientific methods	7(0.10)	2(0.03)	1(0.01)	0(0.00)	0(0.00)	0(0.00)	10(0.14)
Electromagnetic phenomenon and law (EPL)	Static electricity, electric fields, magnetic fields and flux, right-hand rule, electromagnetic induction, electromotive force, Maxwell electromagnetic fields	9(0.13)	3(0.04)	2(0.03)	0(0.00)	0(0.00)	0(0.00)	14(0.19)
Electromagnetic technology and society development (ETSD)	Generator, motor; telecommunication; electromagnetic wave; temperature sensor; relationship between science, technology, and society	4(0.06)	1(0.01)	0(0.00)	0(0.00)	2(0.03)	0(0.00)	7(0.10)

(continued)

Table 10.2 (continued)

Topic	Sub-topic	Remember	Understand	Apply	Analyze	Evaluate	Create	Subtotal
Electrical apparatus at home and application (EAHA)	Types of domestic electrical apparatus, applications of circuits at home, electrical safety at home, electrical safety knowledge, saving electricity	2(0.03)	1(0.01)	2(0.03)	1(0.01)	0(0.00)	0(0.00)	6(0.08)
Subtotal		33(0.46)	22(0.31)	8(0.11)	7(0.10)	2(0.03)	0(0.00)	72(1.00)

Note. The frequency of content statements is presented for each cell. The numbers in parentheses are the proportion of statements in the respective cell to the total number of statements coded

Table 10.3 Physics content standards (grades 7–9) by topic and by cognitive level

Topic	Sub-topic	Remember	Understand	Apply	Analyze	Evaluate	Create	Subtotal
Properties of matter (POM)	Physical properties and change of matter; structure of matter, scale, and application of new materials	17(0.13)	11(0.09)	3(0.02)	1(0.01)	1(0.01)	0(0.00)	33(0.26)
Motion and forces (M&F)	Variety forms of motion, force and mechanical movement	12(0.09)	8(0.06)	9(0.07)	0(0.00)	0(0.00)	0(0.00)	29(0.22)
Electricity (ELE)	Static electricity, electric circuits, magnetism, and effects of interacting fields	3(0.02)	5(0.04)	5(0.04)	0(0.00)	0(0.00)	0(0.00)	13(0.10)
Waves	Light, sound, wavelength, frequency, the speed of waves	12(0.09)	6(0.05)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	18(0.14)
Energy	Types of energy, work, the conversion and transfer of energy, use of energy sources and sustainable development	17(0.13)	16(0.13)	1(0.01)	1(0.01)	0(0.00)	0(0.00)	35(0.28)
Subtotal		61(0.46)	46(0.37)	18(0.14)	2(0.02)	1(0.01)	0(0.00)	128(1.00)

Note. The frequency of content statements is presented for each cell. The numbers in parentheses are the proportion of statements in the respective cell to the total number of statements coded tests at secondary school gateways among New York, United States; Singapore; and Jiangsu, China (Liu et al. 2009). In the 2008 study, Liang and Yuan adopted a method similar to the SEC model (Porter 2002) and analyzed the topics in the Syllabus and the tests. It was found that both the Syllabus and the standardized exams mostly emphasized student learning outcomes at the "understand" cognitive level. Furthermore, the two exams consistently overrepresented the Syllabus at both application and analysis cognitive levels. The study also indicated that neither the organization of the Syllabus nor the exit assessments encouraged creativity, critical thinking, or the development of students' abilities to conduct scientific inquiry.

Since 2008, more studies, normally conducted by college professors and their graduate students, have been conducted to examine alignment between the new secondary school curriculum standards and the gateway exams at the ninth- or the twelfth-grade levels, by applying either Porter's method as used by the SEC or Webb's model (Chen et al. 2010; Guo et al. 2010; Jiao and Chen 2010; Wang et al. 2010). In Chemistry, for instance, one of the most relevant alignment studies was conducted by Wang et al. (2010). Wang and colleagues (2010) examined alignment of the ninth grade exit tests in nine selected provinces and/or cities, covering provincial- and/or municipal-level exams in economically well-developed and underdeveloped regions. In some places, the chemistry content was tested in a single exam, while in other places, the chemistry content was combined with physics and biology in one comprehensive test. With a modified SEC type of approach, the content of chemistry was classified into seven main topics: Solution, States of Matter, Composition of Matter, Structure of Matter, Chemical Changes, Energy, and Chemistry Terminology. The concepts and subtopics within each main topic were then categorized according to the cognitive demands: (1) memorizing, (2) understanding, (3) simple or direct application, and (4) complex application and transfer. For the nine tests sampled, the Porter alignment indices range from 0.59 to 0.74. It appeared that better alignment was associated with the regions with higher economic development levels. However, the researchers did not report the statistical significance of their results. Compared to the municipal-level tests, the provincial-level exams seemed to be of higher quality and better aligned with the curriculum standards. Furthermore, the alignment was higher in a single chemistry exam than that in the combined science test. The identified problems with certain exams included an overly heavy emphasis on "remember" or complex computation in certain tests. Finally, the alignment findings seem to be independent of the types of textbooks in place (Wang et al. 2010).

10.3 Purpose of the Current Study

Over the years, physics courses and exams in China have been perceived as the most difficult of all by many students. We believe that alignment research on standardized physics exams will facilitate the conversation about implementation of standards-based curriculum reform and improvement of learning, instruction, and assessment in physics.

Building on the existing literature and our initial findings in an analysis of physics curriculum standards and the 2009 ninth grade exit physics test in one province (Chen et al. 2010), we launched the current research project by examining the physics exit exams at the secondary schools in several purposely selected geographic locations and by tracking the exams in one place for multiple years (2007–2011).

The research questions are as follows: (1) To what extent are the physics exit exams at the senior high school level aligned with the corresponding curriculum standards (Grades 10–12) in the selected provinces? (2) To what extent are the ninth grade physics exit exams aligned with the corresponding physics curriculum standards (Grades 7–9) in the selected cities?

10.4 Methods

In this section, we first provide some background information about the exit exams at secondary schools in China. We then describe the procedure of data analysis and calculation of the alignment indices. Given that the Porter's method could potentially make the results comparable across different standards documents, assessments, and other components of the educational system, we decided to adopt the Porter's model with a minor modification. We replaced Porter's five categories of cognitive demand with the revised Bloom's taxonomy, i.e., remember, understand, apply, analyze, evaluate, and create (Anderson and Krathwohl 2001) because of the similarity between the two classification schemes. Furthermore, the cognitive domain terminology described in Bloom's taxonomy or the revised Bloom's taxonomy has been widely adopted by Chinese education communities.

10.4.1 The Exit Exams at the Secondary School Gateways in Selected Regions

10.4.1.1 Physics Exit Exams at the Senior High School Level

In most provinces, students must pass a set of exit exams near the end of their senior high school year to be qualified to register for the National College Entrance Examinations in June. Each year, the senior high school exit exams are created by a group of expert high school teachers led by the educational testing agency within each province. Specific testing requirements and the number of subjects tested vary from one province to another. For instance, in provinces such as Guangdong, Jiangsu, and Hainan, physics exit exams are required for liberal arts students only, whereas in other provinces such as Shandong and Ningxia, physics exit exams are required for all students.

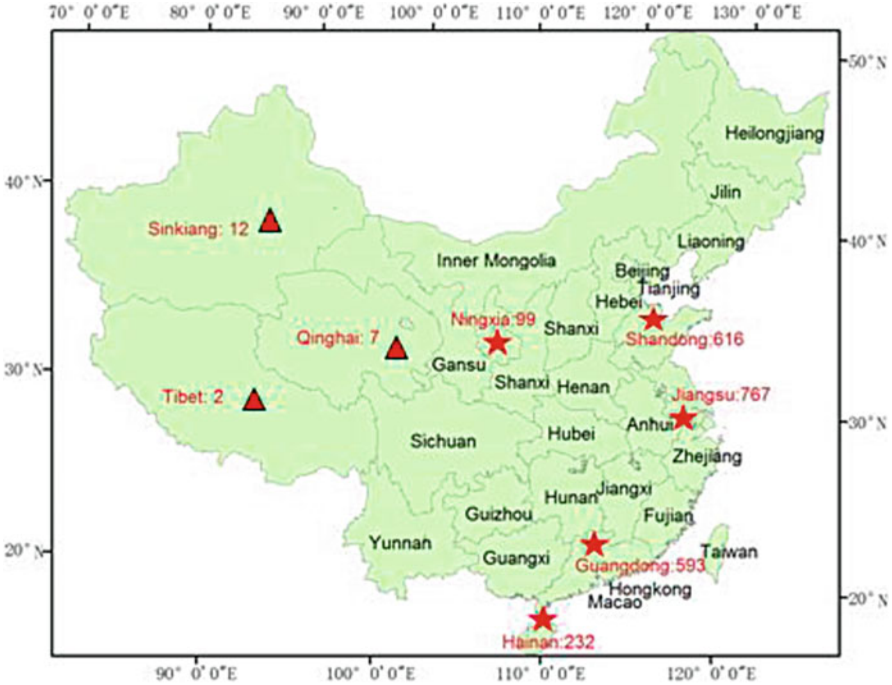

Fig. 10.1 Geographic locations of the sample provinces. Note: Locations labeled with stars represent the five samples. The number next to the name of each individual province represents its population density (number of people per square kilometer). The western China regions labeled with triangles are not included in this study. The population densities in these regions are much lower than those of the selected samples

In this study, we examined physics exit exams from five provinces, ranging from the economically well-developed to the underdeveloped areas in the country: Guangdong (well developed), Jiangsu, Shandong, Hainan, and Ningxia (underdeveloped). These provinces are selected to represent areas of different levels of economic and educational development in China (marked with stars on the map in Fig. 10.1). They are also among the most populated regions in China, representing the first and the second curriculum reform pilot cohorts in the nation. The evaluation of the new school curriculum in the selected provinces would provide evidence and/or lessons for the nationwide implementation of the curriculum reforms.

For the Jiangsu sample, we also examined the exit tests for 5 years (2007–2011) to identify potential trends and patterns. Table 10.2 presents some major characteristics of the physics exit exams in the selected regions.

10.4.1.2 Physics Exit Exams at Junior High School Level

As part of compulsory education in China, junior secondary school students (Grades 7–9) are required to take science courses either as separate or integrated subjects. In most provinces, students take biology in seventh grade, physics in eighth and ninth grades, and chemistry in ninth grade. All students must take the exit examinations before they graduate from the junior secondary school. In Nanjing city, for instance, the ninth grade exit exams for junior high schools are offered in June each year and have seven subjects, including Chinese (120 pts.), mathematics (120 pts.), English (120 pts.), physics (100 pts.), and chemistry (80 pts.). The biology exam is given at the end of eighth grade. Physics exam test items include multiple choice questions, fill-in-the-blank questions, problem-solving questions involving mathematical calculations, and graphing questions or questions related to inquiry-oriented experimental design. Additional required performance assessments (i.e., experiments) are conducted separately by individual teachers on a pass/fail basis. The exit exams serve two functions: first, the test scores gauge the level of student learning against the new curriculum standards and, second, students' total test scores determine the degree of prestige of the senior secondary school to which they can be admitted. Obviously, the ninth grade exit exams are the most important examinations for the students at the compulsory education stage. Parents, teachers, and school principals, as well as the students themselves, all take them very seriously.

According to Wang et al. (2010), the exams tend to have higher quality and better alignment with the corresponding curriculum standards in more economically developed regions. In this study, we purposely selected the Jiangsu Province ninth grade physics exit exams in four cities in 2010, including Nanjing, Changzhou, Suzhou, and Wuxi. Taking Nanjing as an example, we also tracked the exit tests for 5 years (2007–2011) to identify possible trends and patterns. The cities selected in Jiangsu represent economically well-developed areas with a history of regional/national educational leadership. Therefore, the findings of the study could provide valuable lessons for the rest of the country.

10.4.2 Analysis of the Curriculum Content Standards and Exams

The fundamental goal of the physics curriculum standards at the secondary school levels is to develop and improve scientific literacy for all students. Physics curriculum standards at both the lower and upper secondary levels share a common framework defined by three dimensions: knowledge and skills; processes and methods; and emotions, attitudes, and values (MOE 2001, 2003, 2011). In this study, we focus on the analysis of the clearly defined "knowledge and skills" dimension only. The statements on the other two dimensions are not assessed in standardized examinations and are too general for any meaningful productive analysis.

The physics curriculum content standards consist of main topics and subtopics (see Tables 10.2 and 10.3). A number of benchmark statements are also listed under each of the subtopics. To further classify the contents based on the level of thinking required, or the cognitive demands, we used the revised Bloom's taxonomy (Anderson and Krathwohl 2001) to define the cognitive levels with the following sample keywords:

- Remember: know, recognize, identify, recall
- Understand: interpret, translate, explain, illustrate
- Apply: use, implement, calculate
- Analyze: differentiate, distinguish
- Evaluate: critique, judge, reflect
- Create: generate, hypothesize, plan, design

For example, the benchmark statement in the physics standards (Grades 7–9), "students will *know* the relationship of wavelength, frequency, and the speed of wave," was mapped to "waves" and "remember" (Table 10.3). Each benchmark statement is weighted equally. Tables 10.2 and 10.3 present the physics curriculum standards (Grades 10–12 and Grades 7–9) in two 2-D matrices, respectively.

For the analysis of the exams, each test item was classified into a cell of the matrix identical to the curriculum standards analysis grid. After the classification of all test items was completed, the total points of the items in each cell were used as the cell value. The cell value was 0 if no corresponding test items were identified.

One physics education professor and four graduate students with physics teaching experience conducted the data analysis. The average inter-coder reliability was 0.90 for content topics and 0.85 for the cognitive demands. The final results were produced by resolving the disagreements through face-to-face discussions among the research team members.

10.4.3 Measurement of Alignment Between the Tests and Corresponding Standards

To determine alignment between curriculum standards and a test, we first created two tables (one for representing the curriculum standards and the other for the test), each using a two-dimensional matrix in which the rows represent topics/themes and columns represent levels of cognitive demand (such as Table 10.2 for the Grades 7–9 curriculum standards). The values in each cell were converted into proportions out of the grand total, indicating the proportion of total content in the standards document (or exam items) that emphasizes that particular combination of topic and cognitive demand. We then calculated an alignment index using the following equation:

$$P = 1 - \frac{\sum_{i=1}^{n} |(Xi - Yi)|}{2}$$

Here, P is the Porter alignment index, ranging from 0 (indicating no alignment) to 1 (indicating perfect alignment), X denotes cell proportions in Table X (e.g., the standardized exam matrix), Y denotes cell proportions in Table Y (e.g., the curriculum standards matrix), n represents the total number of cells, and i refers to a specific cell in each matrix. For instance, for the 8 × 6 matrix in our study (Table 10.1), $n = 48$. After calculating the alignment indices, we further tested the results for statistical significance by examining the estimated critical values derived from a simulation study (for a more detailed description of this method, see Fulmer 2011). It was found that the critical value was 0.67 for the 8 × 6 matrix and 0.78 for the matrix of five rows and six columns, with the statistical significance at the 0.05 level.

10.5 Results

10.5.1 Alignment of the Physics Content Standards and the Exit Exams at the Senior High School Level

Table 10.4 presents the alignment indices of the exit exams of five provinces in 2010 and the indices of the Jiangsu province exams over 5 years at the senior high school level. Given that the critical value is 0.67, none of the indices are statistically significant at the 0.05 level. However, the numbers are quite close or consistent across the regions and/or across time. The alignment index of the 2010 exam in Jiangsu (representing an economically well-developed region) is almost the same as the one in Ningxia (representing an economically underdeveloped region). This suggests that the test alignment results in a region are not necessarily related to the region's level of economic development. In addition, the test alignment indices in Jiangsu province range from 0.42 to 0.47 during 2007–2011, which indicates that

Table 10.4 The Porter alignment indices (P) of Jiangsu in 2007–2011 and the other four provinces in 2010 (senior high school exit physics exams)

Year	Guangdong	Jiangsu	Shandong	Hainan	Ningxia
2007	n.a.	0.43	n.a.	n.a.	n.a.
2008	n.a.	0.43	n.a.	n.a.	n.a.
2009	n.a.	0.44	n.a.	n.a.	n.a.
2010	0.54	0.42	0.55	0.55	0.41
2011	n.a.	0.47	n.a.	n.a.	n.a.

Note: The critical value is 0.67 with the statistical significance at the 0.05 level

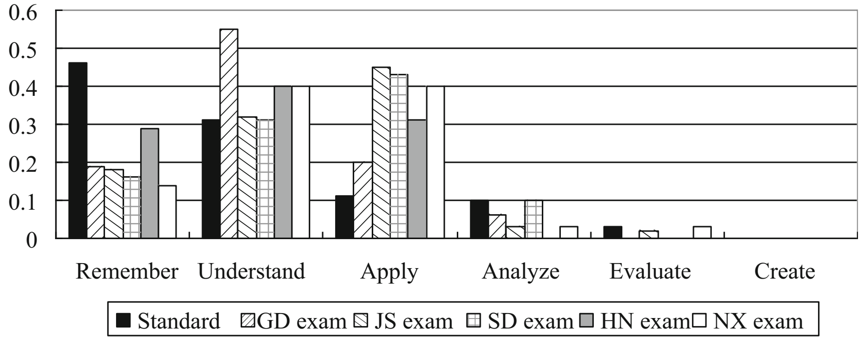

Fig. 10.2 Comparison between content standards (Grades 10–12) and exams in five provinces by cognitive level in 2010

the 5-year implementation of the national curriculum standards did not impact the stability or quality of the exit exams.

For further analysis, we examined the cognitive demands and content coverage separately. Figure 10.2 presents the comparison of alignment analysis across the five provinces by cognitive demands. Compared to physics content standards, the exam items overrepresented the "understand and apply" levels while underrepresented the "remember" level. In the curriculum standards (the black bars in Fig. 10.2), about 46 % of the content is required at the "remember" level and about 11 % at the "apply" level. However, in the five selected exit exams, only about 14–29 % of the test items are located at the "remember" level, while 31–55 % of the content is located at "understand" and 20–45 % at "apply" levels. Across the regions, the proportions of test items classified as "analyze" or "evaluate" seem to be consistently lower than what is required in the standards. Some provinces had excluded items at those two higher cognitive levels in testing. For instance, no items at the levels of analysis and evaluation were found in Hainan's test, while the tests from Guangdong, Shandong, and Hainan did not include any items at "evaluate" level.

In terms of the content coverage of the exams across the five provinces, it was found that most topics required in the standards were covered in the exams. Two exams (from Guangdong and Shandong) covered all main topics in the standards. However, the content distribution in the test did not seem to correlate with the level of emphasis of individual topics as described in the standards. The two topics consistently overrepresented in all five exams are "describing matter" and "interaction forces and laws." The consistently underrepresented topics in the exams include "classical mechanics achievement and limitation" (CMAL), "electromagnetic technology and society development" (ETSD), and "electrical apparatus at home and application" (EAHA). The last two topics (ETSD and EAHA) were not addressed in the exams from Ningxia and Jiangsu (Fig. 10.3).

Taking the Jiangsu exams in five consecutive years as examples, the distribution patterns of the cognitive demands and content are illustrated in Figs. 10.4 and 10.5,

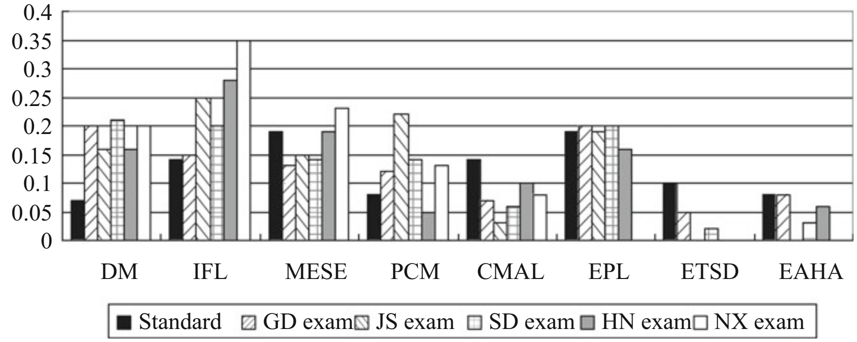

Fig. 10.3 Comparison between content standards (Grades 10–12) and exams in five provinces by topics in 2010

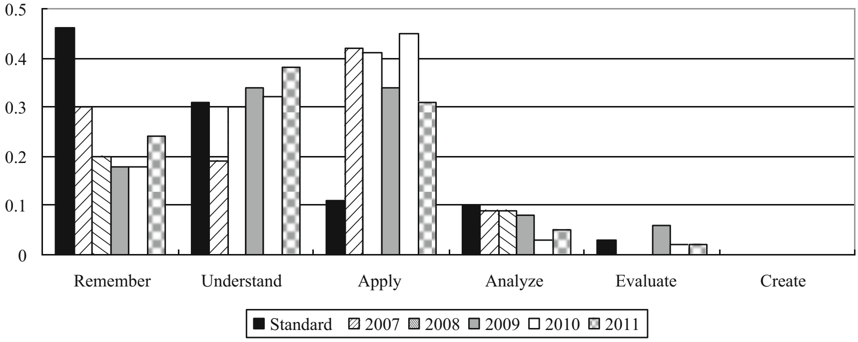

Fig. 10.4 Comparison between content standards (Grades 10–12) and exams for 5 years by cognitive level

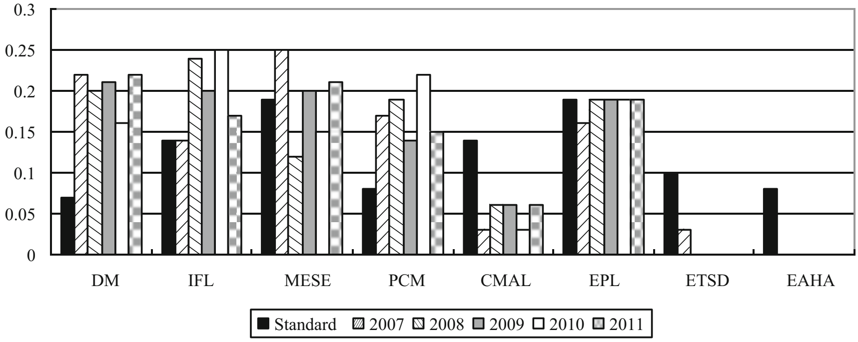

Fig. 10.5 Comparison between content standards (Grades 10–12) and exams for 5 years by topic

respectively. For the distribution of cognitive levels, on one hand, the percentage of the test items at the memory level over the 5 years ranged from 18 % to 30 %, compared to 46 % in the standards. On the other hand, about 31–45 % of the test items were classified as the application level, while the requirement in the standards is 11 %. Apparently, the test items in the Jiangsu provincial exit exams substantially overrepresent the standards in higher cognitive levels (such as application) while underrepresenting standards at memory levels.

In terms of content coverage, compared to the standards, the motion topics related to "describing motion" (DM) and "projectile and circular motion" (PCM) are consistently overrepresented in all exams, whereas the topic "classical mechanics achievement and limitation" (CMAL) is consistently underrepresented. From 2008 to 2011, both "electromagnetic technology and society development" (ETSD) and "electrical apparatus at home and application" (EAHA) topics were not tested.

10.5.2 *Alignment of the Physics Content Standards and Exams at the Junior High School Level*

Table 10.5 presents the alignment indices for the exams from the four selected cities in 2010 and for the Nanjing exams from 2007 to 2011. All indices are lower than the critical value of 0.78, indicating that none of the alignment results is statistically significant at 0.05. However, the alignment indices (ranging from 0.52 to 0.60) are quite stable across the time and the cities within Jiangsu Province.

Figure 10.6 shows misalignment of emphases between the physics curriculum standards and the content distribution in the exams at cognitive demands. Compared to the content standards, the test analysis results demonstrated an apparent shift toward higher cognitive skills by de-emphasizing "remember" and overemphasizing "apply" and "analyze." For instance, in the physics content standards, about 46 % of the content is required at the "remember" level and 14 % at the "apply" level. By contrast, about 30–40 % of points were coded as "apply" and 11–24 % of points as "remember" in the exams. At the "analyze" level, the test items representation was 6–9 %, while the standards requirement is 2 %.

Table 10.5 The Porter alignment indices (P) of Nanjing in 2007–2011 and the other three cities in Jiangsu Province in 2010, ninth grade exit physics exams

Year	Nanjing (NJ)	Wuxi (WX)	Changzhou (CZ)	Suzhou (SZ)
2007	0.55	n.a.	n.a.	n.a.
2008	0.53	n.a.	n.a.	n.a.
2009	0.52	n.a.	n.a.	n.a.
2010	0.54	0.58	0.60	0.55
2011	0.54	n.a.	n.a.	n.a.

Note: The critical value is 0.78

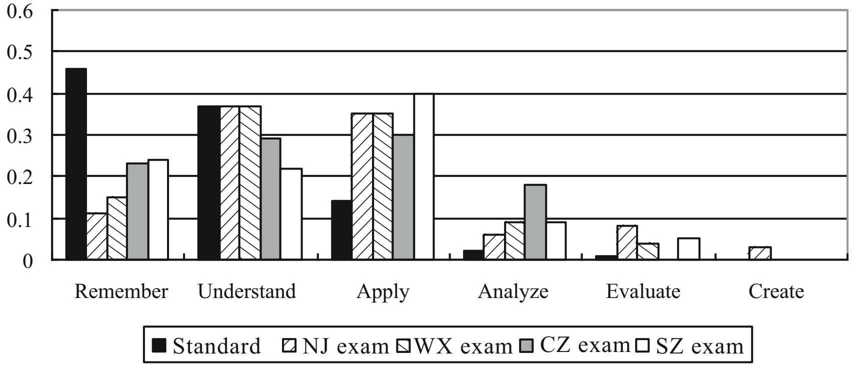

Fig. 10.6 Comparison between content standards (Grades 7–9) and exams in four cities by cognitive demands in 2010

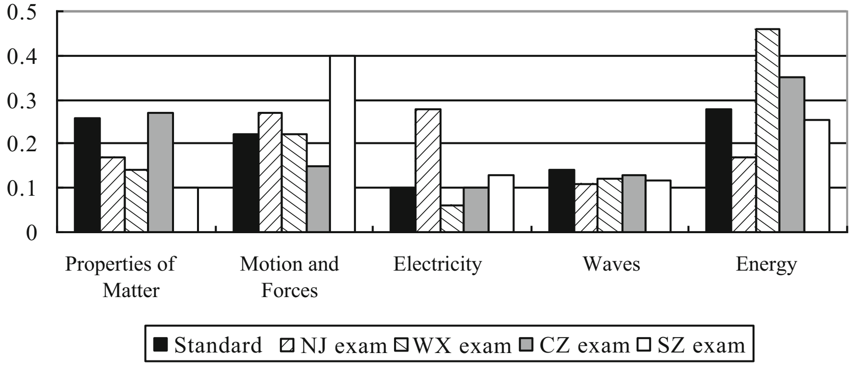

Fig. 10.7 Comparison between content standards (Grades 7–9) and exams in four cities by topics in 2010

The content distribution patterns among the four selected exams are presented in Fig. 10.7. Apparently, the content coverage varies from one place to another. For instance, the most emphasized topics in the standards include "energy" (28 %), "properties of matter" (26 %), and "motion and forces" (22 %), while the least emphasized topic is "electricity" (10 %). However, in the exams, the most emphasized area was "electricity" (28 %) for Nanjing, "energy" (46 %) for Wuxi, and "motion and forces" (40 %) for Suzhou in 2010. Interestingly, the topic on waves was consistently represented at a similar percentage across all four cities.

Following the Nanjing exams as examples, Figs. 10.8 and 10.9 present the overall distribution of cognitive reasoning skills and content coverage as measured by the test items over a 5-year period. Compared to the curriculum standards, the exams consistently de-emphasized memorization of contents and overemphasized higher reasoning skills such as "apply" and "analyze." As for the content coverage, it appears that the topic "electricity" has been consistently overemphasized in the

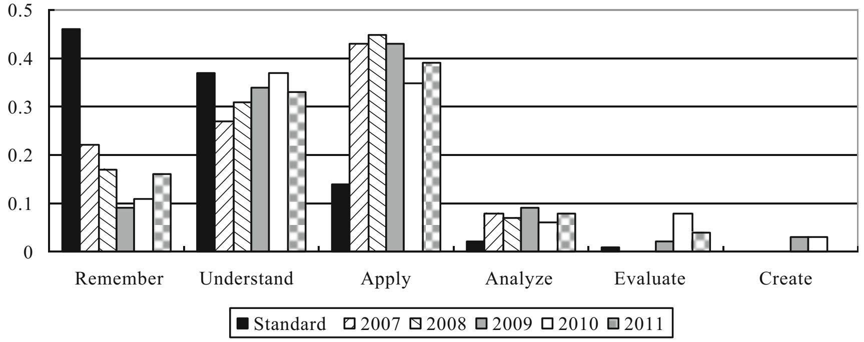

Fig. 10.8 Comparison between content standards (Grades 7–9) and exams over 5 years by cognitive demands

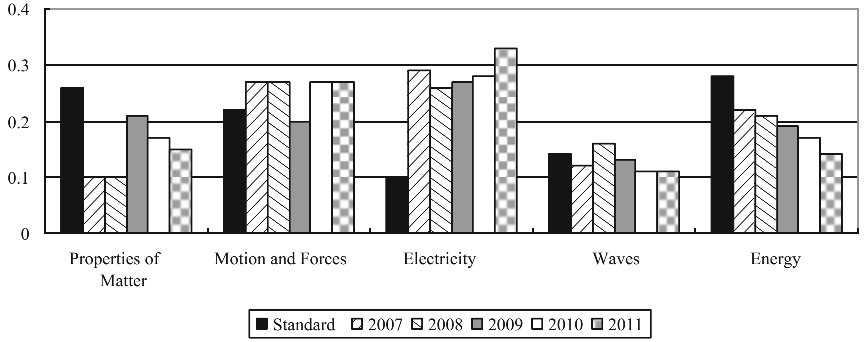

Fig. 10.9 Comparison between content standards (Grades 7–9) and exams over 5 years by topic

exams while other topics such as "properties of matter" and "energy" have been consistently underrepresented.

10.6 Discussion, Implications, and Recommendations

The findings of the current study can be summarized as follows:

1. The alignment indices between the standards and the exit exams at the secondary school level are quite stable for the 5-year window studied in the selected cities or provinces. For the Jiangsu provincial-level high school physics exit exams, the Porter indices range from 0.42 to 0.47, and for the Nanjing city-wide ninth grade exit exams, the Porter indices range from 0.52 to 0.55.
2. The alignment indices across regions and time are all below the corresponding critical values. In other words, none of the alignment results are statistically

significant at 0.05, indicating a general misalignment between the standards and existing physics exit exams

3. Unlike what was found in Wang's and her colleagues' (2010) report, the alignment indices in our study do not seem to be related to the level of economic development of varying geographic locations. Economically well-developed provinces do not necessarily produce physics assessments that are better aligned with the corresponding national curriculum standards
4. Compared to the requirements in the standards, the exams consistently underemphasized "remember" while overemphasizing "apply" and "analyze." Misalignment also occurred when the exams over- or underemphasized content relative to its proportion in the corresponding standards to a varying degree from one year to another.

What can we conclude based on the findings presented above? To what extent should the standardized exit exams at gateways be aligned with the respective curriculum standards? Such questions have not been completely answered here. For instance, by definition, the ninth grade exit exams are criterion-referenced benchmark tests. Students are required to meet the minimum benchmark standards for graduation. Meanwhile, the test scores are also used to "select" students into senior high schools with varying level of prestige. Therefore, the standardized exams are also of a norm-referenced nature. It is reasonable for an exam with a "selective" or "competitive" nature to place emphasis on the higher cognitive demands than those required in the standards. But where do we draw the line between "reasonable" and "unreasonable" levels? The Porter model does not specify the criteria for the alignment index to represent acceptable alignment. In the current study, we used the estimated critical values to determine the level of statistical significance of the results. However, we do not believe that such assessment issues can be solved by statistics alone. In China, the exit exam developers (normally expert teachers) meet and design tests each year based on their own understanding of the curriculum standards as well as the feedback received from teachers, students, and parents. Their knowledge, hard work, and dedication have been reflected in the consistency among the multiple-year alignment indices. We believe that alignment research can be used as a tool to help identify potential or existing problems and provide insights toward the establishment of a more valid and consistent assessment system. The misalignment between the standards and the exams may be addressed in two ways: by reexamining the appropriateness of content coverage and distribution in the curriculum standards and by improving the quality of test items. The test developers may consider constructing a test framework or test blueprint aligned with the respective standards. In addition, there have been concerns that standardized tests are not accurate measures of student achievement in the science education community. Given that a paper-and-pencil test may not be an effective tool for assessing student learning associated with higher levels of cognitive demands (e.g., "evaluate" and "create"), performance assessment or authentic assessment tools (e.g., students' written reports of inquiry projects, experimentation, etc.) should be considered in addition to the traditional paper-and-pencil tests.

The lack of alignment between the curriculum standards and high-stakes tests as identified in this study may influence teaching in both positive and negative directions. The positive aspect might include an emphasis on developing higher cognitive skills in instruction by teachers. During our analysis of the exams, we have noticed some encouraging trends. For instance, more test items involve real-life situations. Some items were designed to assess students' problem-solving and creative thinking skills related to the real world. This will certainly lead teachers to emphasize those higher-order thinking skills in teaching. On the other hand, the overrepresentation of "understand" and "apply" cognitive demands in the exit exams has led to a cognitive overload for many students caused by over-drill or practicing exam-type problems both inside and outside of the school. Students may have achieved high marks in the narrowly defined core subjects at the expense of student learning in other non-tested subject areas. Such misalignment may contribute to the failure in the development of positive attitudes toward school science, student creativity, and their abilities to conduct scientific inquiry in many schools. A substantial number of students have developed high levels of anxiety about the external exams that directly impact both their mental and physical health.

In the following sections, we discuss some limitations of the current study and provide several suggestions or recommendations for future research. First, our alignment research is based on the assumption that curriculum standards and the textbooks are valid and aligned with each other. In our analyses, the main topics and subtopics described in the curriculum standards are classified as general and key concepts or contents. In teaching, the amount of instructional coverage time devoted to each topic is different depending on the classification of the content (e.g., "know," "apply," etc.). We did not assign weight based on the instructional coverage time devoted to different topics, which may have contributed to the "misalignment" results. For future analyses, researchers may want to consider incorporating the suggested instructional time on various topics provided in the curriculum guides and textbooks companions (e.g., Teachers' Guides).

Second, policy researchers have argued that alignment of standards, curriculum, and assessment is the key to supporting implementation of standards-based reform efforts. In this chapter, we focused on the investigation of alignment between high-stakes exams and the corresponding curriculum standards in physics. Such alignment is necessary but not sufficient for successful standards-based curriculum implementation. In reality, the alignment between the standards and classroom instruction can never be assumed. Therefore, we suggest that future alignment studies examine the level of implementation of instructional strategies promoted in the standards. Whereas various approaches such as classroom observations and interviews can be used as fidelity measures, we think that using student/teacher surveys of instructional activities (Fulmer and Liang 2013) might be a more efficient way to get a bigger picture of what's happening at classroom levels in a region. Such surveys can provide information on the content of instruction, level of challenge in instruction, and/or teachers' use of inquiry-oriented, standards-based instructional strategies.

Third, one of the key features of the SEC tools is the use of a common language framework in describing the content areas. In application, however, for various reasons, Chinese researchers tended to modify the SEC framework or adopt a different content classification scheme for data analysis. This creates difficulty in comparing the results across the research studies. We suggest that science education researchers develop a common framework for content analysis based on the Chinese national curriculum standards documents, as a collaborative effort to improve the alignment of learning, teaching, and assessment. In addition, as suggested by other researchers (Martone and Sireci 2009; Polikoff et al. 2011), analyzing the same data with different alignment models (e.g., Webb's model) may provide us with additional insights into agreement and disagreement between standards and assessments.

Finally, based on our review of the alignment studies in China, the data collection and analyses were all done by college professors and their graduate students. There is little evidence that policy makers, test developers, educational administrators at city or provincial levels, and teachers are aware of any of the abovementioned research literature. We therefore suggest that alignment studies be integrated into professional development involving teachers as well as curriculum supervisors in the region.

For future research, we also suggest that different types of experts such as test developers be invited for participation. Different types of experts may view the content and rate exam items differently in terms of the depth of knowledge and the dimensionality of the items (Buckendahl et al. 2000). Involving multiple stakeholders in the alignment research will enhance the credibility of the findings and promote its dissemination to the broader education community more effectively.

It is our hope that our alignment research may prompt the Chinese science educators to reexamine the validity and quality issues related to curriculum standards, textbooks, classroom instruction, and standardized assessment. In addition, it is critical to engage teachers in regular professional development activities related to the above issues. We believe that it is the concerted efforts of stakeholders (e.g., science educator, teachers, educational policymakers, test makers, etc.) that will determine the level of success of the new curriculum reform in China.

Acknowledgment The authors thank Dr. Guo Yuying for her help provided during the preparation of this chapter. The authors also thank Drs. Xiufeng Liu and Gavin Fulmer for their insightful comments and many graduate students in Nanjing Normal University including Jia Cancan, Chen Nin, She Yafang, and Chen Hongli for their assistance in data collection and analysis. The population density and the map of China were provided by Drs. Liyun Lu and Changchun Huang. The authors appreciate it greatly.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Beck, M. D. (2007). Review and other views: "Alignment" as a psychometric issue. *Applied Measurement in Education, 20*(1), 127–135. doi:10.1207/s15324818ame2001_7.

Bhola, D. S., Impara, J.C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice* (fall), *22*, 21–29.

Buckendahl, C. W., Plake, B. S., Impara, J. C., & Irwin, P. M. (2000). *Alignment of standardized achievement tests to state content standards: A comparison of publishers' and teachers' perspectives*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Chen, X., Jia, C. C., & Chen, N. (2010). Alignment between the physics content standard and the exam. *Curriculum, Teaching Material and Method* (课程·教材·教法), *30*(7), 67–71 (In Chinese).

Council of Chief State School Officers. (2004). *Surveys of enacted curriculum*. Madison: Wisconsin Center for Education Research.

Fulmer, G. W. (2011). Estimating critical values for strength of alignment among curriculum, assessments, and instruction. *Journal of Educational and Behavioral Statistics, 36*(3), 381–402.

Fulmer, G. W., & Liang, L. L. (2013). Measuring model-based high school science instruction: Development and application of a student survey. *Journal of Science Education and Technology, 22*(1), 37–46. doi:10.1007/s10956-012-9374-z.

Guo, Y. Y., Xing, T., Xu, G. Q., & Zheng, C. Y. *Alignment between physics curriculum standard and physics achievement qualifying examination of senior high school in China*. 2010 NARST annual international conference.

Jiao, & Chen (2010). Alignment between the biology curriculum standards and the 2009 college entrance biology exams in selected curriculum reform experimental sites. *Measurement and Evaluation in Education* (教育测量与评价), *3*(5), 54–58 (In Chinese).

Liang, L. L., & Yuan, H. Q. (2008). Examining the alignment of Chinese national physics curriculum guidelines and 12th-grade exit examinations: A case study. *International Journal of Science Education, 30*, 1823–1835.

Liu, X., Zhang, B. H., Liang, L. L., Fulmer, G., Kim, B., & Yuan, H. Q. (2009). Alignment between the physics content standard and standardized test: A comparison among US-NY, Singapore, and China-Jiangsu. *Science Education, 93*, 777–797.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 79*(4), 1332–1361.

Ministry of Education (MOE). (2001). *National physics curriculum standards for the full-time compulsory education (grades 7–9). (trial version)*. (初中物理课程标准(实验稿)). Beijing: Beijing Normal University Press (In Chinese).

Ministry of Education (MOE). (2003). *National physics curriculum standards for senior secondary school* (Grades 10-12). (trial version). (高中物理课程标准(实验稿)). Beijing: People's Education Press (In Chinese).

Ministry of Education (MOE). (2011). *National physics curriculum standards for the full-time compulsory education* (Grades 7-9). (义务教育物理课程标准). Beijing: Beijing Normal University Press (In Chinese).

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal, 48* (4), 965–995. doi:10.3102/0002831211410684.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Wang, L., Huang, Q., & Liu, D. F. (2010) Alignment between the chemical curriculum standard (Gr.7-9) and the exams: A case study on chemical concepts. *Basic Education Curriculum* (基础教育课程), *3*, 75–80 (In Chinese).

Webb, N. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education Research Monograph No.6). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7–25.

Yuan, Y. K., Zhao Z., & Yu, Z. Q. (2002). *Commentate on science curriculum standard for compulsory education* (7th–9th grade.) (p. 159) (全日制义务教育科学(7–9年级)课程标准(实验稿)解读). Wuhan: Hubei Education Press (In Chinese).

Zhao, Y. (2009). *Catching up or leading the way: American education in the age of globalization*. Alexandria: ASCD.