An International Perspective on Sixth Grade Mathematics Final Exams between Beijing and New York

Ye Sun, West Virginia University
Shuhua An, California State University, Long Beach
Zhonghe Wu, National University
Xia Li, City University of New York

Abstract

This study compared sixth grade final exams between Beijing and New York in terms of the distribution of types of questions and mathematics content of questions according to NCTM content standards. We used qualitative methods for a document analysis of secondary data sources. Some commonalities, as well as differences, were revealed from these two exams. We found that both exams covered number and operation, algebra, and measurement in their sixth grade mathematics final exams. We also found that the New York exam covered more mathematic topics than the Beijing exam. The Beijing exam focused heavily on number and operations; around 75.5% questions were related to number and operations. In terms of distribution of question types, we found that 71.4% of the New York final exam was multiple-choice questions. Only 18.3% of the Beijing exam was multiple-choice questions. The Beijing exam incorporated more open-ended questions, which demanded more mathematical thinking skills. Possible reasons that contributed to the differences were also discussed.

Introduction

International comparison studies show that Chinese students outperform U.S. students in mathematics (McKnight, et al., 1987; Schmidt, McKnight, Valverde, Houang, & Wiley, 1997; Valverde, Bianchi, Wolfe, Schmidt, Houang, 2002). While various reasons have been investigated, for example, teachers' knowledge (Ma, 1999), and curriculum quality (Li, 2002), very few researchers have looked into assessment, especially the types of questions and the content of questions in final exams. Since assessment is an important measure through which effectiveness of teaching and learning is evaluated (Bransford, Brown, & Cocking, 2000), it is vital to investigate the differences in assessment in international comparison studies.

Since the No Child Left Behind (NCLB) Act, high-stakes testing in mathematics has become more popular in the U.S. The NCLB Act has advocated for annual testing on mathematics and reading in public schools. In Mainland China, high-stakes tests have been implemented for several decades (An, 2000; Li, 2001). Chinese students must pass a highly selective college entrance exam in order to continue onto higher education. It would be very interesting and valuable to examine the content as well as the topics covered in high-stakes tests in different countries.

The body of literature on curriculum and instruction suggests that tests, quizzes, and exams are very important in terms of providing feedback to teachers, thus they play a critical role in the assessment system (Van de Walle, Karp, & Bay-Williams, 2010). Tests, quizzes, and exams provide indicators of student learning performance and the focus of mathematics curriculum. Good exams identify learning objectives or expected learning outcomes, assess mathematical topics to be covered, and provide directions on the learning activities in which students will engage. Exams typically contain information about curriculum goal, curriculum structure, principles of the organization of curriculum content, process of curriculum implementation, and the mechanism of curriculum assessment (Li, 2001). However, knowledge is best assessed with completion, short-answer, and selected-responses items (McMillan, 2007).

Our interests and experiences led us to explore the commonalities, as well as differences, between the Beijing and New York public school mathematics final exams. We would like to investigate the following research questions in the current study:

- How are the types of exam questions distributed in these two final exams?
- How are the mathematics topics distributed in these two final exams?

Review of the Literature

Comparison Studies between China and the U.S.

Cross-national studies indicate that U.S. students learned many more mathematics topics than other countries in a given academic year. U.S. curriculum is characterized as “a mile wide and an inch deep” (Valverde, et al., 2002). However, Chinese curriculum is criticized for being in the opposite direction. It is very traditional and places more emphasis on operations. In addition, it covers fewer math topics than other countries (Liu & Sun, 2002).

In recent years, comparison studies between China and U.S. started to capture scholars’ interests in the field of mathematics education. For example, Cai (1995) carried out a study to investigate problem-solving skills between U.S. and Chinese students. He found that Chinese students did better on computations of number and operations than their U.S. counterparts. Sun and Liu (2001) studied two sets of exams from U.S. and China; they found that Chinese exams focused more on topics in number and operations, while U.S. curriculum focused more on measurement and real world applications. Li (2002) examined the differences of the integer addition and subtraction representations between two curricula and pointed out the substantial differences between them. Li’s study also agrees with previous findings that Chinese curricula focused more on number and operations using rational numbers, while U.S. curriculum tended to pay less attention to the topic. In particular, some US curricula totally ignore the operations involving rational numbers. Li’s study also reports that neither of the curricula emphasized problem-solving skills.

Classroom Assessment in Mathematics

National Council of Teachers of Mathematics [NCTM] published *Assessment Standards for School Mathematics* in 1995, in which it defined assessment as: “the process of gathering evidence about a student’s knowledge of, ability to use, and disposition toward mathematics and of making inferences from that evidence for a variety of purposes” (NCTM, 1995, p.3). Many scholars agreed that classroom assessment goes beyond high-stakes testing. For instance, alternative assessment such as: classroom observation, interview, and portfolio could also be used as classroom assessment strategies (NCTM, 1995; Wu & An, 2007). However, other scholars argued that tests, quizzes and exams are still the dominant type of assessment used in classrooms (Hopkins, 1998; McMillan, 2007). In particular, 77% of the grades are reported from the tests, quizzes and exams (Senk, Beckman, & Thompson, 1997). Furthermore, studies tended to indicate that exams are valuable in terms of identifying a student’s weaknesses in learning mathematics. Thus, teachers could adjust their teaching focus to accommodate the students’ needs (Glaser & Silver, 1994).

Large-scale assessment in mathematics education, such as Trends in Mathematics and Science Study (TIMSS) (2003) and Program for International Student Achievement (PISA) (2006), reported some common mathematics topics that were covered in these exams. For example, number and operations, and problem-solving skills are the major objectives in these large-scale international comparison studies. They also reported that some mathematics topics (e.g., statistics) are taught at an earlier age in some countries, while in other countries, they are not covered (Gal, et al., 1999; McKnight, et al., 1987; Schmidt, et al., 1997; Telford, M. & Caygill, 2007; Valverde, et al., 2002).

Multiple-choice, short-answer, and open-ended questions are the most common types of questions found in tests and exams (Abrams & Madaus, 2003; Eckstein & Noah, 1993). The multiple-choice questions are defined as questions that ask students to identify the correct answer from a given set of choices. Short-answer questions are defined as questions that have a definite answer and use some words or procedures to show the answer. These types of questions often require the final answer, including a brief demonstration of related mathematical thinking processes. Open-ended questions are defined as questions that do not have any fixed procedure to follow, and students have more freedom to solve the problems. The open-ended questions usually have more cognitive demands or thinking processes required in solving a problem than the other two types of questions. Open-ended questions often require students to justify their mathematics thinking process in detail. Researchers continue to report that exam's format and exam's content are connected. They also report a growth in trend of using multiple-choice questions to substitute the traditional short-constructed questions (Eckstein & Noah, 1993). The functions of exams and tests are two-fold. They can help to provide feedback to teachers to inform classroom instructions. They also provide information on what kind of knowledge has been successfully obtained by the students. However, with more emphasis on high-stakes testing, there is also a danger that teachers would only cover the topics in tests, and ignore any other topics that are not covered in tests, thus "teaching to the tests" (Abrams & Madaus, 2003).

Theoretical Framework

The researchers have examined the body of literature to define the theoretical framework for this study. The NCTM five content standards for school mathematics including: number and operation, algebra, geometry, measurement, probability and statistics (NCTM, 2000) have provided a direction for the content analysis. Meanwhile, the American Association of Advancement of Science [AAAS] has also proposed a six-category analysis framework including: number concept, number skills, geometry concept, geometry skills, algebra graph concepts, and algebra equation concept (AAAS, 1993), which is included in the NCTM five category analysis framework.

We decided to use five criteria proposed by NCTM study in our current research for the following reasons. First, after close examination of the criteria proposed by the researchers mentioned in previous sections (the review of literature), we noticed that they tend to agree that number, algebra, and geometry are common topics. Another reason is that the content areas in the new standards in Chinese Mathematics Education are aligned with the NCTM standards (Liu, & Sun, 2002). So, we only used NCTM criteria in the current study.

In addition, we used item types discussed by various researchers to analyze the types of questions in our study. According to Haladyna (1997), open-ended items are more appropriate than selection items such as multiple-choice because open-ended items measure high-inference mental skills or abilities. The short-answer items can "assess thinking skills when students are required to supply a brief response to a question or situation that can be understood only by the use of the targeted thinking skills" (McMillan, 2007, p.203). Multiple-choice items are typically not favored for measuring higher order skills (Hollingworth, Beard, & Proctor, 2007).

Methodology

Selection of Documents

Four criteria were used to select exam documents from Beijing and New York.

First, we chose Beijing and New York because these two cities are similar in both size and economy. Second, these exams would be considered as final exams in the public schools in each city. Third, these exams would be considered to be used at the 6th grade level. Fourth, these exams were published around the same time. Thus, we chose two final exams from each city; one is from Beijing (2005), and the other one is from New York (2005). Both documents were analyzed in the original language and were examined for adherence to NCTM standards (NCTM, 2000) and assessment standards (NCTM, 1996). We used qualitative methods for a document analysis of secondary data sources (Miles & Huberman, 1994; Patton, 1990).

Data Collection

Data were collected from the following two sources:

- Sample sixth grade exam paper of 2005 compiled from the New York State Education Department.
- Sixth grade exam paper of 2005 compiled by Beijing public schools.

Data Analysis Procedures

First, we identified the learning objectives of each exam by looking at each of their standards. Then, we studied each exam carefully using NCTM content framework: number and operations, algebra, geometry, measurement, probability and statistics. Two bilingual teacher-educators (Chinese and English), independently read and coded the documents. Qualitative methods were applied to examine these two exams.

The final exams chosen for this study were analyzed in the original language. Two of the authors analyzed the exams in content areas using the analysis framework that was proposed by NCTM and analyzed types of questions defined by McMillan (2007).

Results

Distribution of Types of the Exam Questions

Three types of questions, multiple-choice, short-answer, and open-ended questions, were found in both the Beijing and New York exams. We found that the New York 6th grade exam had many more multiple-choice questions than their counterparts in Beijing (71.4% vs. 18.3%). Table 1 shows the distribution of types of test questions. The Beijing 6th grade exam had higher percentages of short-answer and open-ended questions than their counterparts in New York. In the analysis of distribution of question types, 41.0% of questions in the Beijing 6th grade final exam were classified as short-answer, while only 28.6% of the New York 6th grade exam questions were similarly classified.. We also noticed that 40.8% of Beijing final exam questions were categorized as open-ended, while none of questions in the New York final exam was categorized as being open-ended..

Table 1

Distribution of Test Question Types

	Multiple-choice		Short-constructed		Open-ended	
Area	Beijing	New York	Beijing	New York	Beijing	New York
Total questions	5	25	24	10	20	0
Percent	18.3%	71.4%	41.0%	28.6%	40.8%	0

It is obvious that the Beijing exam included more open-ended questions than New York. Open-ended questions have a higher level of cognitive demand. For example, question 49 in the Beijing exam is described in the following:

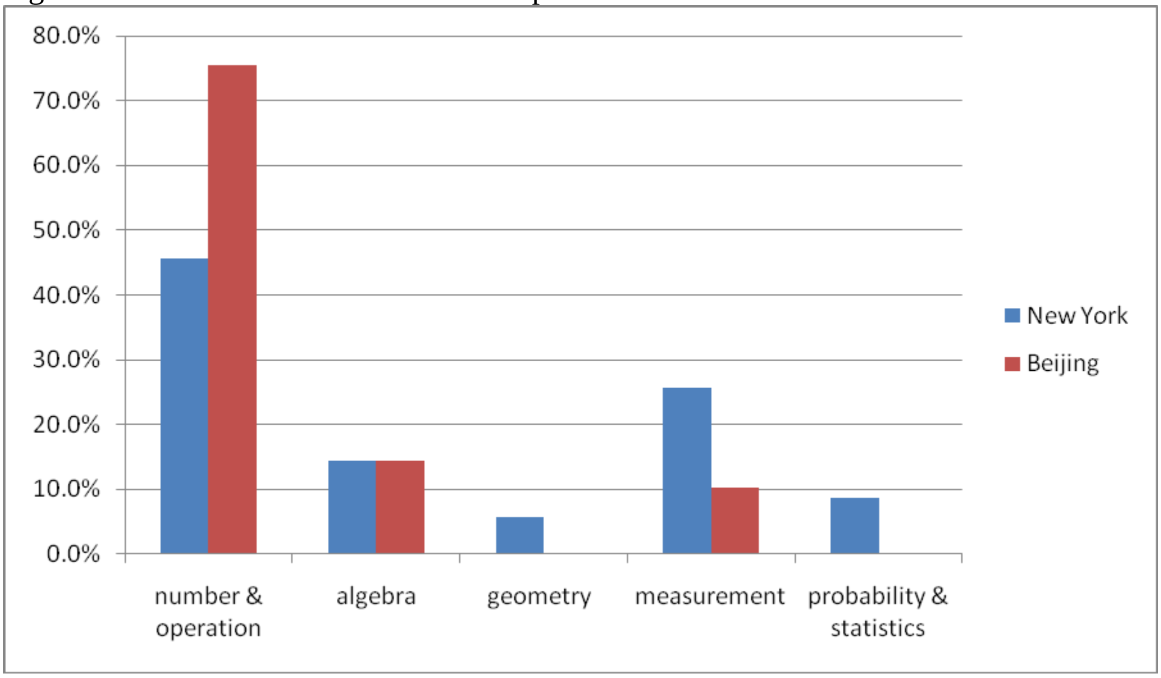
There are two logs with the length of 30 dm and 80 dm. We would *like to cut these two logs into pieces that have the same length (the length should be integer number). We should not have any logs left over. What is the length of each piece of log?*

This question indicates that Beijing students were required to demonstrate the clear procedures (including reasons) when solving a mathematics problem. They are often encouraged to write out the procedure, especially the mathematics procedures and their thinking process from the instruction of the exam.

Distribution of Mathematics Topics

We found some similarities, as well as differences, in terms of the number of mathematics topics covered in each exam (see Figure 1). Both the New York and Beijing exams covered number and operations, algebra, and measurement concepts. However, the scope of mathematics topics spanned all five content standards in the New York exam. Only three mathematics topics were covered in Beijing 6th grade exam.

Figure 1. Distribution of Mathematics Topics



More specifically, 75.5% of the questions in the Beijing exam focused on number and operations, while only 45.7% of the questions in the New York exam focused on number and operations. It seemed that both New York and Beijing devoted the same amount of questions to algebra content (14.3%). We also found that 25.7% of the questions in the New York exam were related to measurement, while only 18.3% of the questions in the Beijing exam pertained to measurement. The percentage of questions covered in the New York exam regarding geometry, probability, and statistics is 5.7% and 8.6% respectively. No questions in the Beijing exam aimed at these topics.

A close study on the item level revealed some differences in representing the following topics: number and operations, algebra, and measurement.

Number and operations. The Beijing exam focused more on operations involving rational numbers. There are few operations on rational numbers in the New York exam. In addition, the Beijing exam also tested students to solve rational number problems mentally. It is worth mentioning that not only the accuracy, but also the speed of solving problems, were required in the curriculum. For example, children must work out 10 to 12 such problems in one minute.

A sample question in number and operations in the Beijing exam:

Mental math: $300*0.3\%=$

A sample question in number and operations in the New York exam:

On Friday and Saturday, there were a total of 200 cars in the parking lot of a movie theater. On Friday, 120 cars were in the parking lot. What percent of the total number of cars were in the parking lot on Friday?

These two examples show that the New York questions related to real-world applications, whereas the Beijing questions focused on computational fluency.

Algebra. The Beijing exam focused on the procedure of solving one-variable equations and understanding the concept of variables, while the New York exam tended to emphasize patterns or using multiple representations to model a real-world situation.

We also found out that questions were usually put in a real-world context and were often represented with graphs and pictures in the New York exam, while most of the Beijing exam questions were related to pure mathematics symbols.

A sample question in the New York exam:

Barry is training to be a gymnast. He increases the number of push-ups each week by following a number pattern. The number of push-ups Barry does for 5 weeks is shown in the table below.

Table 2

Barry's push-ups

Week	Number of push-ups
1	16
2	19
3	22
4	25
5	28

If Barry continues to do push-ups according to the number pattern, how many push-ups will he do during the 10^{th} week? Anne predicts that Barry will do 59 push-ups during the 15^{th} week. On the lines below, use words, symbols, or numbers to explain whether Anne's predication is correct.

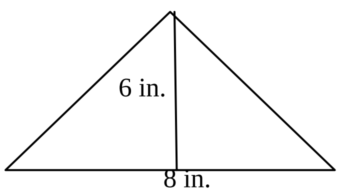
A sample question in the Beijing exam: *Solve the following equation: 5-x=2*

These two examples show that the New York questions used tables to ask students to find out patterns in order to solve the algebra questions, while the Beijing questions asked students to solve one-variable equations using mathematics symbols.

Measurement. The Beijing exam tended to focus on conversion skills among different measurement units, while the New York exam placed more emphasis on using formulas to determine the area of a certain region. We also found out that the Beijing exam required students to remember the formulas, while in the New York exam, the formulas were given in the questions. So, there is no need for students to remember the formulas.

A sample question in the New York exam:

Willard has a stained glass window with one triangular piece, as shown below What is the area, in square inches, of the triangular piece?



$A = \frac{1}{2} bh$

A sample question in Beijing exam:

Please draw a circle with a diameter of 4 cm, and calculate the circumference.

These two examples indicate that the New York exam used a real-world context and visual representations to present the measurement question, while the Beijing exam asked students to figure out the circumference of the circle using paper and pencil.

Discussion

The analysis of the exam documents between the two cities in this study suggests that there are some similarities as well as differences. Three types of questions: multiple-choice, short-answer, and open-ended questions were found in both the Beijing and New York exams. Both exams covered number and operations, algebra, and measurement concepts. We also found that the New York exam covered all five mathematics content standards, while the Beijing exam only covered three mathematics content standards. The results agree with the findings from international studies that U.S. curricula covered more topics than other countries (Gal, et al., 1999; McKnight et al., 1987; Schmidt, et al., 1997; Valverde, et al., 2002). A closer investigation of the mathematics topics revealed that the Beijing exam required a lot more on number and operations, especially operations involving rational numbers at the 6th grade level. Mental math was also a focus area in the Beijing exam. This is also in line with the previous discoveries (Cai, 1995; Sun, & Liu, 2001; Li, 2002). A possible reason for these differences might be that Chinese curriculum is very mathematical-oriented. Since topics like geometry, probability, and statistics are focused on less in the 6th grade mathematics in China, they were not covered at all in the 6th grade mathematics exam. In contrast, the U.S. curriculum is very broad and is arranged in a spiral manner. It includes every mathematics topics, in each grade, for different purposes.

We also found that the New York exam used a lot of multiple representations in algebra and usually provided a real-world context when a mathematics problem was presented. However, it was less focused on the basic knowledge and basic skills such as number and operations. The Beijing exam was just the opposite. It emphasized the students' grasp of the basic knowledge and basic skills, for example, mental math. These differences in content coverage might be due to different standards and curricula. We realized that the 6th grade is considered a middle school level in most US states, and it is still at the elementary level in China. The differences in the two exams might also be due to the beliefs in the education systems. The Chinese educational system is a top down system with very few variations across different provinces. Mathematics coherence (including basic knowledge and basic skills in number and operations) is important in mathematics education in China. Teachers and curriculum developers have fewer concerns on the real world applications (or multiple representations) as much as the coherence of the mathematics system. Although the Chinese mathematics education reform de-emphasized the requirement of basic knowledge and basic skills in public schools mathematics curriculum, it takes time for mathematics educators to understand it and apply it in the final exam.

Another difference discussed in the current study is the distribution of the question types. The majority of questions in the New York exam were multiple-choice, while the majority in the Beijing exams were short-answer and open-ended questions that would require students to explain their answers. Therefore, the Beijing exam shows more cognitive demands than the New York exam. Since language is an instrument to express thinking, if students could use languages or math symbols to justify their answers, they would demonstrate mathematics understanding more deeply. The New York exams reflect the findings reported by Eckstein and Noah (1993) that multiple-choice questions replace the short-answer questions. In contrast, the Beijing exam does not reflect the trend reported by Eckstein and Noah (1993). The different distributions in the two exams may be due to the consideration of expenses associated with time, money, and reliable scoring in the U.S. (Hollingworth, Beard, & Proctor, 2007). However, there is a consistent belief among researchers that open-ended items are able to measure high-order thinking (Haladyna, 1997).

One of the limitations of this study is that only one city and one grade of exams were analyzed. Thus, the findings of this study may not be generalizable to all cities/states in both countries. Further studies are needed to investigate multiple grade levels with a variety of regions in both countries.

The results of this study might be applied to the classroom practice. U.S. teachers might want to incorporate more materials on number and operations in teaching and testing. Chinese teachers might want to incorporate more real-world examples and applications into classroom teaching and testing.

References

Abrams, J. & Madaus, G. (2003). The lessons of high-stakes testing. *Educational Researcher, 33*(1), 4-14.

American Association for the Advancement of Science. (1993). *Benchmark for Science Literacy*. New York: Oxford University Press.

An, S. (2000). Globalization of education in China. *International Journal of Education Reform, 9*(2), 128-133.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn.* Washington D.C.: National Academy Press.

Cai, J. (1995). A cogitative analysis of U.S. and Chinese students' mathematical performance on tasks involving computation, simple problem-solving and complex problem-solving. *Journal for Research in Mathematics Education monograph series*, Reston, VA: National Council of Teachers of Mathematics.

Eckstein, M.A., & Noah, H.J. (1993). *Secondary school examinations: International perspectives on policy and practice*. New Haven, CT: Yale University Press.

Gal, I., Van Groenestijn, M., Schmitt, M.J., & Tot, D. (1999). Numeracy framework for the international adult literacy and life skills survey (ALL). Ottawa, Canada: Statistics Canada and Washington, DC: National Center for Education Statistics.

Glaser, R., & Silver, E. A. (1994). Assessment, testing and instruction: Retrospect and prospect. In L. Darling-Hannond (Ed.), *Review of research in education* (pp. 393-419). Washington, D.C: American Education Research Association.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Boston, MA: Allyn and Bacon.

Hollingworth, L. B., Jonathan J., & Proctor, T. P. (2007). An investigation of item type in a standards-based assessment. *Practical Assessment Research & Evaluation. 12*(18), 1-13.

Hopkins, K. (1998). *Educational and psychological measurement and evaluation*. Boston, MA: Allyn and Bacon.

Li, Y. (2002). A comparison of integer addition and subtraction problems presented in American and Chinese Mathematics Exams. In Sowder. J & Schappelle, B. (Eds.), *Lessons learned from Research National Council of Teachers of Mathematics*. Reston, VA: National Council of Mathematics.

Li, Z. (2001). History of Chinese high-stakes testing. Unpublished thesis. Bejing Normal

University**.**
Liu, J., & Sun, X. (2002). *Understanding mathematics curriculum standards.* Beijing, China: Beijing Normal University Press.
Ma, L. (1999). *Knowing and teaching elementary mathematics: Teaching's understanding of fundamental mathematics in China and United States*. Mahwah, NJ: Erlbaum.
McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes
McMillan, J. H. (2007). *Classroom assessment.* Boston, MA: Pearson Education, Inc.
Miles, M. B., & Huberman, M. A. (1994). *An expanded qualitative data analysis* (2^{nd} ed.) ThoU.S.nd Oaks: Sage.
National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: NCTM.
Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2^{nd} ed.). Newbury Park, CA: Sage.
Schmidt, W.H., McKnight, C.E., Valverde,G.A., Houang, R.T., &Wiley, D.E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics* (Vol. 1) Dordrecht, The Netherlands: Kluwer.
Senk, S. L., Beckman, C.E., & Thompson, D.R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187-215.
Sun, Y., & Liu, J. (2001). Comparative study on the mathematics examination at grade four in Beijing and New York. *Proceedings of the International Conference on Primary School Science and Math Education*, Beijing, China.
Telford, M., & Caygill, R. (2007). PISA 2006: How ready are our 15-year-olds for tomorrow's world? [online report] Retrieved from http://www.educationcounts. govt.nz/publications/ series/2543/pisa_ 2006/16616
Valverde, G. A., Bianchi, L..J., Wolfe, R.G., Schmidt, W.H., Houang, R.T. (2002). *Using TIMSS to investigate the translation of policy into practices through the world of textbooks*. Dordrecht, The Netherlands: Kluwer.
Van de Walle, J. A., Karp, K.S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally.* Boston: Allyn & Bacon.
Wu, Z., & An., S. (2007). Alternative assessment in mathematics teaching. In N. Noraini (Ed.), *Classroom assessment in mathematics education*. New York: McGraw-Hill Education.