CrossMark

Teachers' perceptions of effective teaching: a comparative study of elementary school teachers from China and the USA

Lingqi Meng[1] · Marco Muñoz[2]

Abstract This international study investigated Chinese and American elementary school teachers' perceptions of effective teaching. The sample comprised Chinese ($n=108$) and US ($n=110$) participating teachers. The Effective Teaching Quality Survey (ETQS) was adopted for this comparative education research, an instrument that operationalized Stronge's effective teaching theoretical framework. Research questions that guided this study were as follows: (1) what are the differences between the USA and China's elementary school teachers' perceptions regarding effective teaching? And (2) when comparing teaching experience, school location, and effective/ineffective teachers, what are the differences between the USA and China's elementary school teachers' perceptions on effective teaching? Statistically significant results were found when comparing nationality, teaching experience, school location, and effective/ineffective teaching. How these results related to the US and Chinese educational contexts with respect teaching and learning are discussed.

✉ Lingqi Meng
13173182557@163.com

Marco Muñoz
marco.munoz@jefferson.kyschools.us

[1] College of Education, Qufu Normal University, Qufu, Shandong 273165, China

[2] Data Management & Research Office, Jefferson County Public Schools, Data Management, Planning, and Program Evaluation VanHoose Education Center, 3332 Newburg Road, Louisville, KY 40218, USA

1 Introduction

Effective teaching has been a hot topic in international education during the last several decades. There are at least two factors that stimulate this research agenda associated with effective teaching. One factor is the large-scale international studies such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). These test results revealed that students in Eastern countries outperformed their counterparts in Western countries. The USA ranked at an average level among all participating countries with respect to student mathematics achievement (Mullis et al. 2003, 2008). Another factor that helped understand this phenomenon is the worldwide educational reform initiated by different countries. Both China and the USA have launched ambitious reform agendas to enhance the quality of K-12 education. A common feature of the educational reform efforts in China and the USA is to seek learning about effective teaching from other cultures. For example, some US researchers (Stigler and Hiebert 1999) claimed that mathematics teaching in Eastern countries was more effective than in the USA and that US teachers can learn about effective teaching approaches from Eastern countries. Numerous articles and books were published to introduce the way of Eastern teaching and learning (e.g., Fan et al. 2004; Ma 1999).

Meanwhile, Eastern countries were eager to learn from the USA to reform their education. An evidence is that the Chinese mathematics standards (CMOE 2001) were largely adopted from US teaching beliefs to characterize what effective teaching should be. From this international perspective, investigating the USA and China's teachers' perceptions of effective teaching is important. The different perceptions may help policy makers and practitioners determine how and to what extent it is helpful to adopt an effective teaching strategy from other countries.

The term effective teaching can be interpreted in different ways. Theoretically speaking, different sets of learning theories have different views on effective teaching. Kirshner (2002) claimed three metaphors for effective teaching: (1) habituation, (2) construction, and (3) enculturation. A habituationist view of effective teaching emphasized on skill acquisition, so teachers can use repetitive practice to make teaching effective. A constructivist view of effective teaching focused on conceptual understanding, so teachers can facilitate students' learning trajectories to help them learn concepts. An enculturalist view of effective teaching is situated learning that emphasized culturally involvement in the classroom teaching, so effective teaching must help students acquire cultural dispositions.

Meanwhile, effective teaching has been continuously explored in school effectiveness research since the 1960s in Western countries (Teddlie and Reynolds 2000). Teddlie and Liu (2008) explored effective teaching in both effective schools and ineffective schools in China and found that teachers from more effective schools performed better on six measures of teaching effectiveness. These six measures of teaching effectiveness are as follows: (1) maintaining an environment conductive to learning, (2) maximization of instructional time, (3) management of learner behaviors, (4) effective delivery of learner behavior, (5) presentation of appropriate content, and (6) providing opportunities for student involvement. In China, effective teaching has been theorized in four ways. The first way is based on economics principles that emphasize investing small and gaining big profit (i.e., high return-on-investment). A

parallel argument in education is that practices that teachers in class devote the least time and effort end up resulting in maximal students' learning. The second way is that effective teaching must be built on evidence that students make real progress, a point adopted from US value-added modeling (Muñoz and Chang 2007; Muñoz et al. 2011). The third way claims that effective teaching must be explored in three levels, the ideological level, the strategic level, and the concrete method level. The fourth way is featured as high-efficient classroom teaching that is in accordance with beliefs and values advocated by Chinese reform standards documents.

To measure effective teaching, researchers (e.g., Devine et al. 2013; Grant et al. 2013; Griffin 2013; Liu and Meng 2009; Muñoz et al. 2013; Stronge 2007, 2010; Stronge et al. 2011; William 2010) have explored empirically based categories and themes from different countries. Among these explorations, Stronge (2007) framework of effective teaching has been well applied and introduced in the USA and in Eastern societies. For example, some states in the USA (e.g., Virginia, New Jersey, Georgia) have adopted this work as their teacher evaluation framework. This framework is also introduced to China along with the book of *East meets West* translated into Chinese (Grant et al. 2014). Stronge's framework has the following six categories as Liu and Meng (2009) described: (1) prerequisites of effective teachers (e.g., verbal ability, knowledge of teaching and learning, certificate status, content knowledge, teaching experience), (2) the teacher as a person (e.g., caring, shows fairness and respect, interactions with students, enthusiasm, motivation, dedication to teaching, reflective practice), (3) classroom management and organization (e.g., classroom management, organization, discipline of students), (4) planning and organizing for instruction (e.g., importance of instruction, time allocation, teachers' expectations, instruction plan), (5) implementing instruction (e.g., instructional strategies, content and expectations, complexity, questioning, student engagement), and (6) monitoring student progress and potential (e.g., homework, monitoring student progress, responding to student needs and abilities). In fact, Stronge's framework was a synthesis of the extant literature in effective teaching. It was derived from 27 works of research-based effective teacher qualities. William (2010) further confirmed that the above six categories were highly connected to the teacher effective research. Stronge (2013) finished a validation report of his framework. For instance, regarding the content validity, Stronge provided details on how his framework was related the previous teacher effective research and state standards. The construct validity was also very good (Cronbach's alpha >0.74 for K-12 school samples).

Although Stronge's framework was developed in the USA, many Chinese educators' reflection of effective teaching can fit in most of this framework. Kan (2013) developed a 22-item instrument regarding effective teaching. Twenty items in Kan's questionnaire have very close conceptualization to Stronge's framework. Wang (2010) talked about effective mathematics teaching as students-teacher co-participation in classroom activities, providing materials interesting to students, using manipulatives, inquiry-oriented teaching, and reflecting their own teaching practice frequently. Wang's work was largely in accordance with Stronge's dimensions (2) the teacher as a person and (5) implementing instruction. Zhang and Deng (2010) investigated Chinese high school students regarding effective teaching in chemistry. They found the following items related to effective teaching: student participating in activities, comfortable learning environment, freely expressing ideas in class, independent thinking, having

peer collaboration, doing chemical experiments, facilitating creative thinking, having a fair assessment policy in class, selecting a good pedagogy, and preparing lessons in accordance with the new curriculum standards. These items were closely related to Stronge's dimensions (4) planning and organizing for instruction and (5) implementing instruction. Liu and Luo (2013) surveyed 928 Chinese high school teachers' view of effective teaching. They found that participants emphasized preparing teaching materials that can cause students' interests, demonstrating kindness to students, frequently checking students' progress, and creating a comfortable learning environment. Stronge's six-dimension framework was also adopted by comparative researchers investigating the USA and/or China's effective teaching. Liu and Meng (2009) used it to study perceptions of Chinese teachers, students, and parents with respect to effective teaching and learning. William (2010) adopted this framework to investigate administrator and teacher perceptions of the qualities of effective teachers. Muñoz et al. (2013) investigated the "black box" of effective teaching in the USA. Meng et al. (2015) investigated Chinese high school teachers' perceptions of effective teaching. In this study, we adopted Stronge's six dimensions as theoretical framework. Although we reviewed a variety of perspectives regarding effective teaching, we focused on teachers' teaching behaviors, teaching skills, and morality to measure effective teaching in this study.

Two research questions for this study were as follows: (1) what are the differences between the USA and China's elementary school teacher perceptions regarding effective teaching? And (2) what are the differences of teaching experience, school location, and highly effective/less effective teaching on the characteristics established by the Effective Teaching Quality Survey (ETQS) between the USA and China's elementary school teachers? Descriptive statistics with measures of central tendency and dispersion were used for answering the first research question. The second research question was answered by the non-parametric Kruskal-Wallis test and the parametric factorial analysis of variance (ANOVA).

2 Method

2.1 Sample

The sample in this study comprised elementary school teachers from China ($n=110$) and the USA ($n=113$). A power analysis was conducted to avoid committing a type II error. According to Olejnik (1984), a one-way ANOVA, with alpha=0.05 level of significance, medium effect size, and statistical power=0.70, needed a sample of 100 participants in each country.

US participants were drawn from 90 elementary schools located in a high-poverty urban district in the Midwest. Chinese participants were selected from six elementary schools in the northeast of China. US response rate was 62.78 % which is considered acceptable for social science research. China's response rate was 94.7 %. Although the two regions in the USA and China are not comparable in terms of their economic development and culture (as it is typical on international studies), both of them were representative (on measures of central tendency) for their own country regarding their economic and K-12 educational contexts. The Chinese urban sample was drawn from a

regional-level city and two county-level cities; furthermore, three schools were located in the rural area and other three schools were located in cities. These schools are average schools in their own region regarding student achievement and they participated in this study on a voluntary basis. Of the six schools selected in China, three were rural schools that can be regarded as low SES schools; however, the other three schools can be regarded as high SES schools. A reason to identify the three rural schools as low SES schools is that many students who have good economic standing in rural areas tend to go to county or city schools in recent years; as result, students who stay in rural schools are poorer than students in urban schools in China. Since all US participants were from an urban school setting, their students' economic conditions in title I schools were similar to China's rural schools. In this study, we categorized school location as low and high socio-economic status (SES) in each country. This classification did not mean that the USA and China had perfect matching samples which never happens in international studies due to differences in educational context. However, students with low SES in both China and the USA shared similar family background (e.g., low family income). Meanwhile, they also had some differences. Students in US low SES schools were ethnic minorities, in contrast to Han ethnicity of Chinese counterparts. Although they were usually identified as poor performing students in their country, there was no research conducted to compare these two groups of students' academic performance. We do not know whether Chinese students in the sampling schools perform higher than the US counterparts. A common perception in the large-scale international studies (e.g., PISA) is that Chinese students performed higher than the US counterparts. "Chinese students" in these international studies were from the most developed areas in China (i.e., Shanghai), not from the region in China that was selected for this study. In our study, the current classification helped us understand the different perceptions between teachers who had low teaching quality and those who had high teaching quality.

Five cases were identified as invalid when reviewing both questionnaires, three of them from the US sample and two from China's sample. Three US invalid questionnaires did not contain enough demographic information. Two Chinese invalid questionnaires were only filled to less than half of the information. As a result, the analytic sample was 218 in total ($n=108$ for China, $n=110$ for the USA). The majority of participants in the US sample were females ($n=104$); however, the Chinese sample was balanced in terms of gender ($n=42$ for males and $n=66$ for females).

2.2 Instrumentation

The questionnaire adopted for this study was the Effective Teaching Quality Survey (ETQS) (William 2010). ETQS has been applied in several worldwide quantitative studies (e.g., Meng et al. 2015; Muñoz et al. 2013) in teacher effectiveness. The survey has four parts. In part I, participants were asked to rank indicators of teacher qualities in five categories illustrated in Stronge's (2007) framework: (1) classroom management and organization, (2) planning for instruction, (3) implementing instruction, (4) monitoring student progress, and (5) teacher as a person (Liu and Meng 2009). In part II, participants were asked to rank the headings of the five categories. Part III was a writing request for any additional indicators of teacher qualities not presented in the

survey. Part IV asked participants to fill out general information including gender, years of teaching experience, education level, ethnicity, and school designation.

Unlike the traditional Likert-type 5-point scale survey, ETQS was a forced-choice questionnaire. For the indicators in each category, respondents were asked to rank the importance among them in a comparable way. That is, if a category contained five indicators, the participants needed to rank them from 1 to 5 in order of their perceived impact on student achievement. Number 1 stood for the strongest perceived impact and number 5 was the lowest perceived impact. Guskey (2007) clearly discussed the advantage of using a forced-choice survey and concluded that comparing similar items can avoid receiving similar rating scores as using a Likert-type scale questionnaire. William (2010) conducted two tests (20 participants for each) and found a validity score of 0.86. In addition, the language critique for the survey was also conducted and a few statements were revised (William 2010, p. 68). In recent years, researchers (e.g., Meng et al. 2015; Muñoz et al. 2013) have applied a forced-choice survey approach to educational comparative research when applicable for purposes of data collection.

In this study, we used two versions of the ETQS (English and Chinese) due to the language difference in the USA and China. The Chinese version of ETQS was obtained from the study of Meng et al. (2015). One indicator "use data to make instructional decisions" embedded in the category of "Monitoring Student Progress" was deleted in the Chinese version since it was not suitable to the Chinese educational context as explained in the study of Meng et al. (2015). We recalculated the rest of indicators in the category of "Monitoring Student Progress" by multiplying 1.25 to the Chinese sample to match the US data. In addition, we deleted the demographic variable of ethnicity in this study since all Chinese participants are from the same ethnicity of Han.

2.3 Data collection and analysis

US data was collected using Internet-based survey techniques (Dillman et al. 2009) from elementary school teachers in an urban district located in the Midwest region. The ways to identify effective teachers involved conducting a two-level hierarchal linear model (HLM), as explained in details in the study of Muñoz et al. (2011). The Chinese data was collected from a northeastern province of China. A Chinese professor from a state-level training center was asked to select three elementary schools in cities and three elementary schools in a rural area. They represented the regular elementary schools in their area. School principals were asked to identify about 20 in-service teachers, with 10 more effective and 10 less effective. Effective teachers in this selection meant that a teacher was capable of increasing students' test scores with excellent teaching skills as perceived by Chinese principals. This was a regular method to identify effective teachers in China. Unlike US school principals, Chinese school principals had detailed information regarding student test scores. They investigated students' progress on a semester basis. As a result, they knew which teacher made students maximal progress, a way we called value-added evaluation. Another different situation in China was that Chinese teachers had a lot of chances (e.g., school level, district/county level, regional level, and state level) to take part in a teaching skill competition. A small number of teachers in each school represented their school to participate in different kinds of competition. So school principals knew these teachers very well in terms of their teaching skills (e.g., excellent oral presentation, well-

organized activities, and teaching innovations). A common perception was that a teacher who had excellent teaching skills will help his/her students maximize their learning (including increasing their test scores). In summary, although we did not use the same method to identify effective teachers in both countries, we argued that effective teachers identified in this study have a common feature: they are capable of helping students make academic progress.

This study involving primary data was reviewed and approved by Qufu Normal University (China). The US data were drawn from a secondary database approved by the University of Louisville and the Jefferson county public schools. No personally identifiable information was collected, no individual data was released, and only aggregated data was reported.

Two graduate students (A and B) were trained for ensuring accuracy and propriety in data collection in China. Graduate student A went to city schools, while graduate student B went to rural schools. They first contacted the school principals to schedule 30 to 40 min with participants, and then they took hardcopies of the questionnaire to the schools. Participants gathered in the same room and could ask questions when filling out the questionnaire. The principals/vice principals and the graduate student entered the data collection room together. The principals/vice principals helped by distributing the questionnaire to participants. The less effective teachers were given a questionnaire with a number I after the questionnaire title; this helped researchers identify more-effective/less-effective teachers, but none of participants knew this difference. Graduate students A and B told participants that this survey was only for research purposes so they were expected to express real opinions, with no response being correct or incorrect. As all participants completed the questionnaire, they turned them-in to the graduate students.

In terms of data analyses, both the Kruskal-Wallis test and the analysis of variance (ANOVA) were used for this study. Since the scale of measurement of the ETQS indicators are ordinal, we first used the Kruskal-Wallis test for data analysis; this non-parametric test that is appropriate for ranked data was followed with an ANOVA since it is considered a robust test (i.e., strong when handling violations to assumptions) and parametric test (Hinkle et al. 2003) and re-assesses findings of the non-parametric counterpart. Four demographic variables served as independent variables, while the indicators that characterized effective teachers were treated as dependent variables. The names of the dependent variables in this study were the same as in the study of Muñoz et al. (2013) study, except for variety 1 and variety 2 in the third category of the ETQS. Two items "Employs a variety of techniques and instructional strategies to accomplish learning goals" and "Uses a variety of questioning techniques" in the study of Muñoz et al. (2013) were labeled as the same name "variety." In our study, we labeled the first one as "Variety 1" and the second one as "Variety 2" separately. Table 1 shows the codes for the four independent variables included in this comparative education study.

For purposes of data analyses and from a statistical power perspective associated with balancing sample cell sizes, the researchers established a minimum sample size of 29 for each level of the independent variables (Hinkle et al. 2003). For example, the variable teaching experience was collapsed due to the unbalanced cell sample sizes. This variable had five options: 1–5, 6–10, 11–15, 16–20, and 20 years above. Accordingly, the cell sample sizes were 48, 50, 39, 21, and 58. Since one of them is below 29, we recoded the third and fourth options together and obtained new balanced

Table 1 Codes of independent variables

Variables	Codes
Gender	1=male, 2=female
Teaching experience	1=between 3 and 9 years, 2=between 10 and 14 years, 3=between 15 and 19 years, 4=20 or more years
School location	1=high socioeconomic status, 2=low socioeconomic status
Teacher effectiveness	1=less effective, 2=highly effective

cell sample sizes: 48, 50, 60, and 58. Finally, the new codes for teaching experience were 1=1–5 years, 2=6–10 years, 3=11–20 years, and 4=20+years. Regarding to levels of education, the Chinese sample contained 5 teachers with a master degree, 76 teachers with a bachelor's degree, 23 teachers with an associate degree, and 2 teachers had missing data. In contrast, the US sample included 31 teachers with post-master's degree of education, 66 teachers with a master's degree, and 13 teachers with a bachelor's degree. However, since there were no matching subsamples with respect to participants' degrees in the two countries, we did not conduct a statistical analysis on the levels of education in this study. Regarding the school location, we categorized low and high socioeconomic status schools as comparable labels. Low socioeconomic status schools in the USA are referred as title I schools, while high socioeconomic schools stood for non-title I schools. In China, poor schools referred to rural schools and rich school referred to city schools. Regarding the gender differences, we only considered China's sample for a statistical analysis since a balanced sample size (42 males vs. 66 females) was obtained. The majority of US teachers were females (104 out of 110); in this country, this is the norm at the elementary school level since most male teachers tend to be found at the secondary school level.

3 Results

3.1 Overall descriptive comparisons between China and the USA

3.1.1 Overall means

Table 2 described the means on each of the items by US and Chinese participants. Since means in a ranking survey were highly relied on the items within the category, it is not comparable between two means in different categories. For example, $M=2.46$ in the first category is about on the middle with a four-item ranking. However, $M=4.62$ is also about on the middle with an eight-item ranking. One cannot compare these two mean values. The following comparisons are within category only. In the first category of classroom management and organization, US and Chinese teachers indicated totally different preferences when comparing the four items. US teachers selected "Order and routines" ($M=2.16$, SD=1.11) as the most important item and "Discipline" as least important one ($M=2.67$, SD=0.98). In contrast, Chinese teachers selected "Physically and emotionally safe environment" as the most important item ($M=1.97$, SD=1.10)

Table 2 Descriptions of variables and means

Labels	Descriptions	Means		
		Total	USA	China
Order and routines	Maintains order and routines	2.26	2.16	2.36
Preparation	Prepares materials ahead of time and has them ready to use	2.69	2.46*	2.93*
Physically and emotionally safe environment	Maintains a physically and emotionally safe environment for students	2.09	2.20	1.97
Discipline	Reinforces expectations for positive behavior and responds to misbehaviors promptly	2.71	2.67	2.74
Limits interruptions and focuses class time	Limits interruptions and focuses classroom time on teaching and learning	3.80	3.65	3.95
High expectations	Establishes and communicates high expectations for student achievement	3.23	2.15*	4.33*
Pacing	Maintains appropriate pacing of instruction	4.38	4.32	4.44
Considers student learning styles	Considers student learning styles and plans instruction accordingly	2.28	2.72*	1.83*
Links instruction to objectives	Links instruction to objectives	3.30	3.06*	3.55*
Links instruction to real life	Links instruction to students' real-life	3.21	3.75*	2.82*
Variety 1	Employs a variety of techniques and instructional strategies to accomplish learning goals	2.46	2.44	2.48
Engage	Designs lessons to actively engage students in the learning process	2.05	1.90	2.19
Variety 2	Uses a variety of questioning techniques	4.06	4.19	3.93
Guided practice	Provides clear examples and offers guided practice	3.39	2.90*	3.88*
Grouping	Uses effective grouping strategies	4.03	4.81*	4.03*
Higher-order skills	Focuses instruction on higher-order skills rather than memorization of information	4.00	3.61*	4.39*
Homework	Uses homework to augment student learning	4.18	4.35*	3.99*
Feedback	Gives clear, specific, and timely feedback	2.35	2.63*	2.06*
Re-teaching	Re-teaches when students do not achieve mastery	2.99	2.35*	3.63*
Assessment	Selects appropriate assessment tools and strategies to evaluate student progress	2.62	2.47	2.77
Concern	Demonstrates concern for the physical and emotional well-being of students	3.73	3.60	3.87
Fairness	Treats all students with fairness	3.60	3.89*	3.30*
Respect	Treats all students with respect	2.96	3.14	2.79
Interaction	Interacts and fosters positive relationships with students	3.57	2.87*	4.29*
Content knowledge	Displays an excitement for subject area content	4.73	4.62	4.84

Table 2 (continued)

Labels	Descriptions	Means		
		Total	USA	China
Excitement	Displays an excitement for teaching and learning	4.33	3.75*	4.93*
Commitment	Demonstrates and on-going commitment to the profession	5.63	6.05*	5.20*
Reflection	Use reflection to improve his or her own practice	5.68	5.13*	6.25*
Teacher as a person	Teacher as a person	2.51	3.24*	1.78*
Classroom management and organization	Classroom management and organization	2.36	2.08*	2.64*
Planning for instruction	Planning for instruction	2.94	2.70*	3.19*
Implementing instruction	Implementing instruction	2.94	2.55*	3.34*
Monitoring student progress	Monitoring student progress	3.72	3.53*	3.93*

*Represents statistically significance in K-W test

and "Preparation" as the least important one (M=2.93, SD=1.13). In the second category of planning for instruction, both US and Chinese teachers selected "Pacing" (M=4.32, SD=1.48 for the USA; M=4.44, SD=1.24 for China) as the least preference among six items. With respect to the most important item, US teachers selected "High expectations" (M=2.15, SD=1.35) in contrast to "Considers student learning styles" (M=1.83, SD=1.12) for Chinese teachers.

In the third category of implementing instruction, both US and Chinese teachers selected "Engage" (M=1.90, SD=1.22 for the USA; M=2.19, SD=1.43 for China) as the important item, while US teachers selected "Grouping" and Chinese teachers selected "Higher-order skills" (M=4.81, SD=1.50 for the USA; M=4.39. SD=0.86 for China) as the least important one. In the fourth category of monitoring student progress, on one hand, both US and Chinese teachers selected "Homework" (M=4.35, SD=1.18 for the USA; M=3.99, SD=1.20 for China) as the least important item; on the other hand, US teachers selected "Re-teaching" (M=2.35, SD=1.16) as the most important item and Chinese teachers took "Feedback" (M=2.06, SD=0.94) as the most important one.

In the fifth category of teacher as a person, US and Chinese teachers have totally different preferences when comparing the eight items. US teachers picked "Interaction" (M=2.87, SD=2.14) as the most important item and "Commitment" (M=6.05, SD=2.58) as the least important one. While Chinese teachers selected "Respect" (M=2.79, SD=1.82) as the most important item and "Reflection" (M=6.25, SD=2.25) as the least important one. In the sixth and last category, which asked teachers to rank the importance of the previous five categories (see last five items in Table 2), teachers in both countries selected "Monitoring student progress" (M=3.53, SD=1.37 for the USA; M=3.93, SD=1.36 for China) as the least important one. US teachers selected "Classroom management & organization" (M=2.08, SD=1.24) as the most important

item while Chinese teachers selected "Teacher as a person" ($M=1.78$, SD=1.45) as the least important one among the five categories.

3.1.2 Statistical differences between the USA and China

A Kruskal-Wallis analysis demonstrated 21 significant findings. After using the Bonferroni adjustment that prevents from committing a type I error and that consists of dividing the 0.05 alpha level by the total number of tests (i.e., 0.001), the significant items reduced to 13. The statistically significant differences at alpha level 0.001 were as follows: high expectations (X^2 (1)=77.00, $p<0.001$), considers student learning styles (X^2 (1)=18.24, $p<0.001$), links instruction to real life (X^2 (1)=21.83, $p<0.001$), guided practice (X^2 (1)=26.56, $p<0.001$), grouping (X^2 (1)=22.31, $p<0.0001$), higher-order skills (X^2 (1)=13.39, $p<0.0001$), re-teaching (X^2 (1)=49.87, $p<0.0001$), interaction (X^2 (1)=30.25, $p<0.001$), excitement (X^2 (1)=16.77, $p<0.001$), reflection (X^2 (1)=16.54, $p<0.001$), teacher as a person (X^2 (1)=39.66, $p<0.001$), classroom management and organization (X^2 (1)=18.90, $p<0.001$), and implementing instruction (X^2 (1)=24.58, $p<0.001$).

The means of the above 13 significant items were listed in Table 2. The smaller means for US teachers were on items of high expectations ($M=2.15$, SD=1.35 vs. $M=4.33$, SD=1.59), guided practice ($M=2.90$, SD=1.43 vs. $M=3.88$, SD=1.27), higher-order skills ($M=3.61$, SD=1.64 vs. $M=4.39$, SD=1.86), re-teaching ($M=2.35$, SD=1.16 vs. $M=3.63$, SD=1.19), interaction ($M=2.87$, SD=2.14 vs. $M=4.29$, SD=1.68), excitement ($M=3.75$, SD=2.07 vs. $M=4.93$, SD=2.00), reflection ($M=5.13$, SD=2.37 vs. $M=6.25$, SD=2.25), classroom management and organization ($M=2.08$, SD=1.24 vs. $M=2.64$, SD=0.97), implementing instruction ($M=2.55$, SD=1.09 vs. $M=3.34$, SD=1.07). These indicated that US teachers emphasized these items more than their Chinese counterparts.

In contrast, the items that Chinese teachers achieved smaller means were as follows: considers student learning styles ($M=1.83$, SD=1.01 vs. $M=2.72$, SD=1.56), links instruction to real life ($M=2.82$, SD=1.413 vs. $M=3.81$, SD=1.588), grouping ($M=4.03$, SD=1.343 vs. $M=4.81$, SD=1.499), teacher as a person ($M=1.78$, SD=1.449 vs. $M=3.24$, SD=1.761). The results illustrated that Chinese teachers emphasized on these items more than US teachers.

In addition, thee items, grouping, higher-order skills, and re-teaching, were more significant than the other ten items since the p values for these three items were less than 0.0001. The practical values of these items will be articulated in the discussion part.

3.2 Significant findings on demographic variables between the USA and China's elementary teachers

3.2.1 School location

The Kruskal-Wallis test on school location further demonstrated that five items were statistically significant. These findings were as follows: considers student learning styles (X^2 (1)=12.29, $p<0.01$), excitement (X^2 (1)=5.93, $p<0.05$), classroom

management and organization (X^2 (1)=5.19, $p<0.05$), planning for instruction (X^2 (1)= 3.83, $p<0.05$), and implementing instruction (X^2 (1)=5.62, $p<0.01$).

A factorial ANOVA test revealed that five items were statistically significant without interactions with nationality. These findings were as follows: limits interruptions and focuses class time [F (1, 218)=4.26, $p<0.05$], high expectations [F (1, 218)=5.60, $p<0.05$], considers student learning styles [F (1, 218)= 14.58, $p<0.01$], excitement [F (1, 218)=5.88, $p<0.01$], and classroom management and organization [F (1, 218)=6.67, $p<0.01$]. The test also showed that two items had statistically significant interactions. For commitment, we found an interaction between country with school location [F (1, 218)=4.69, $p<0.05$]. For implementing instruction, we found an interaction between country with school location [F (1, 218)=5.53, $p<0.05$].

In sum, three items, considers student learning styles, excitement, and classroom management and organization, were statistically significant in the above two tests. Teachers in low socioeconomic schools ranked "Considers student learning styles" and "Classroom management and organization" as more important items than teachers in high socioeconomic schools. By contrast, teachers in high socioeconomic schools ranked "Excitement" as a more important item than teachers in low socioeconomic schools.

3.2.2 Comparisons between highly effective and less effective teaching

The Kruskal-Wallis test indicated only one item, pacing, was statistically significant (X^2 (1)=6.44, $p<0.01$). On the other hand, four items were statistically significant when using a factorial ANOVA test: preparation [F (1, 217)=4.34, $p<0.05$], pacing [F (3, 216)=6.74, $p<0.01$], re-teach [F (1, 218)=5.57, $p<0.01$], and grouping [F (1, 218)=4.43, $p<0.05$]. No item had an interaction with nationality in this test. The mean for less effective teachers was 4.15, in contrast to the mean of 4.61 for highly effective teachers (see Table 3). We concluded that less effective teachers emphasized more on maintaining appropriate pacing of instruction than highly effective teachers.

Table 3 Means of school location and effectiveness for significant dependent variables

Labels	School location		Effectiveness	
	Poor (n=125)	Rich (n=93)	Less (n=108)	Highly n=110
Planning for instruction (1=highest priority to 4=lowest priority)				
Pacing	(4.42)	(4.33)	4.15	4.61
Considers student learning styles	1.98	2.68	(2.29)	(2.27)
Teacher as a person (1=highest priority to 8=lowest priority)				
Excitement	4.90	4.49	(4.72)	(4.74)
Teacher qualities (1=highest priority to 5=lowest priority)				
Classroom management and organization	2.19	2.58	(2.39)	(2.33)

The means in parentheses are not significant

3.2.3 Teaching experience comparisons

The Kruskal-Wallis test showed three significant items: re-teach (X^2 (3)=9.17, $p<0.05$), assessment (X^2 (3)=11.93, $p<0.01$), and interaction (X^2 (3)=8.18, $p<0.05$). Meanwhile, the factorial ANOVA analysis showed that only one item "Assessment" was statistically significant on teaching experience [F (3, 216)=3.914, $p<0.01$]. Three items had significant interactions: preparation [F (3, 216)=2.690, $p<0.05$], high expectations [F (3, 216)=3.25, $p<0.05$], and concern [F (3, 216)=2.835, $p<0.05$]. It was noted that assessment is the one item both statistically significant in the two tests. Finally, a post hoc analysis was conducted to reveal the details, which found that the significant difference was only between teachers with 6- to 10-year teaching experience [M=2.10, SD=1.13] and teachers with 20+-year teaching experience [M=2.97, SD= 1.35] (see Table 4). We concluded that teachers with 6- to 10-year teaching experience emphasized more on selecting appropriate assessment tools and strategies to evaluate student progress than teachers with 20+-year teaching experience.

3.3 Significant findings on gender differences in the Chinese sample

The Kruskal-Wallis test showed four significant items: discipline (X^2 (1)=4.00, $p<0.05$), considers student learning style (X^2 (1)=5.73, $p<0.05$), variety 1 (X^2 (1)= 8.12, $p<0.01$), and guided practice (X^2 (1)=3.91, $p<0.05$). On the other hand, the ANOVA test showed the above four items were statistically significant on gender: discipline [F (1, 108)=4.28, $p<0.05$], considers student learning style [F (1, 108)= 6.36, $p<0.01$], variety 1 [F (1, 108)=6.57, $p<0.01$], and guided practice [F (1, 108)= 4.20, $p<0.05$]. These results were interpreted as follows. On the one hand, male teachers in China ranked three items "Discipline," "Variety 1," and "Guided practice" as more important than female teachers in China. On the other hand,

Table 4 Means of teaching experience and gender for significant dependent variables

Labels	Gender (China)		Teaching experience (China and the USA)			
	Male (n=42)	Female (n=66)	1–5 years (n=48)	6–10 years (n=50)	11–20 years (n=60)	20+years (n=58)
Classroom management and organization (1=highest priority to 4=lowest priority)						
Discipline	2.50	2.89	(2.67)	(2.72)	(2.68)	(2.76)
Planning for instruction (1=highest priority to 6=lowest priority)						
Considers student learning style	2.17	1.62	(2.23)	(2.24)	(2.28)	(2.38)
Implementing instruction (1=highest priority to 6=lowest priority)						
Variety 1	2.95	2.18	(2.55)	(2.53)	(2.35)	(2.40)
Guided practice	3.57	4.08	(3.43)	(3.27)	(3.18)	(3.62)
Monitoring student progress (1=highest priority to 4=lowest priority)						
Assessment	(2.62)	(2.86)	2.66	2.10	2.68	2.97

The means in parentheses are not significant

female teachers in China ranked "Considers student learning style" as a more important item than male teachers (see Table 4 for detailed mean values).

4 Discussion

4.1 US and China's elementary school teachers' perceptions of teacher effectiveness

The main purpose of this study was to compare elementary school teachers' perceptions of teacher effectiveness between the USA and China. Both similarities and differences were found when comparing total means in each category of the survey and using the Kruskal-Wallis test, followed by complementary ANOVAs. These results reflected different cultural and educational contexts in each country. We discussed these findings in the following paragraphs.

When comparing the total means in each category of the teacher effectiveness framework that guided this comparative education study, similarities were found between both countries on the perceptions of teacher effectiveness of elementary school teachers. Both the USA and China's elementary school teachers perceived "pacing," "homework," and "monitoring student progress" as the least priority while indicating the value of "engage" as the highest priority within their category. These findings may be explained as a result of the influence of the educational contexts in the two countries.

For "pacing," we argue that the delivery of lessons at the elementary school is not as demanding as middle schools' or high schools' teaching processes. Although pacing allows schools to avoid unnecessary repetition and makes sequencing of content a more rational activity (by providing support for vertical alignment), the process is relatively simpler at the elementary school level than at the secondary school level. Therefore, elementary school teachers may not pay much attention on maintaining appropriate pacing of instruction. This was evidenced in the study of Meng et al. (2015) where high school teachers did not select "pacing" as the least important indicator. In the USA, pacing is always seen as an approximate schedule that lays out how much time it usually takes to cover each set of lessons of each unit of study (Saphier et al. 2008). However, veteran teachers do not feel as pressured as novice teachers to follow the pacing guides. In that sense, pacing guides are typically more useful for novice teachers who have not taught the curriculum in the past. An additional complexity associated with pacing in the USA is that schools operating in different socio-economic context might need to decelerate or accelerate depending on the prior knowledge brought by the students at the beginning of the school year. US teachers are likely to feel free to use their professional judgment when deciding about their pacing because of the emphasis in the NCLB legislation that teaching is about student performance—making sure student learn the content is what the teaching job is about. It is no longer admissible for teachers to think that teaching the content as planned on the pacing guide is their job and that it is the student's responsibility to learn it. NCLB emphasized that no child should be left behind with a gap in knowledge.

For "homework," we argue that US students do not get assigned as much homework at the elementary school level, so teachers did not value homework as critically important as other in-school indicators. Chinese teachers always assign students more

homework when compared to other countries, so there is little space to give students even more homework to improve their learning in China. Meanwhile, the fact that Chinese teachers did not value homework as an important indicator may also relate to the emphasis of high-efficiency teaching in China (e.g., Long 2010). High-efficiency teaching asks teachers to design lesson plans and to manage teaching time efficiently. As a result, like US teachers, Chinese teachers may ignore indicators that do not happen inside the classroom as really valuable for effective teaching.

The finding of the USA and China's teachers selecting "monitoring student progress" as the least priority indicated that teachers do not pay much attention to the current wave of educational reform policy that promotes formative assessment (Black and Dylan 2009). US and Chinese teacher training programs needed to help elementary school teachers understand that effective teaching should include monitoring students' progress. This finding is in contrast with the findings in the study of Liu and Luo (2013) where Chinese high school teachers perceived "monitoring student progress" as a very important factor to effective teaching.

The finding of "engage" as the highest priority in both the USA and China is consistent with the study of Meng et al. (2015). The authors explained that teachers in the two countries perceived engaging students in the learning process as critically important due to the influence of the new curriculum standards, the Common Core State Standards (CCSS) in the USA and the New Curriculum Standards (CMOE 2001) in China. These curriculum standards expected effective teaching to engage students in teaching activities. It is critical to maximize engaged instructional time for all students since it is definitely a precursor to academic learning time. This finding is also consistent with the study of Grant et al. (2013) where both the USA and China's award-winning teachers perceived "having high student engagement" is important (p. 251). This finding seems to support a long pattern of importance for engagement in the US education (e.g., Connor et al. 2010; Marzano 2007). However, it is only in the last decade that engagement has been emphasized in China.

Three different perceptions of effective teaching were found in the first dimension of classroom management and organization, the fourth dimension of monitoring student progress, and the sixth dimension of teacher qualities of the survey regarding the total means. This may imply different values embedded in teachers' thinking. For the US teachers, the selection of (a) order and routines, (b) re-teaching, and (c) classroom management and organization as the most important indicators reflect behaviorist beliefs of teaching and learning as well as the need to address the uniqueness of the high-poverty urban context in the USA. Behaviorism has impacted US teaching and learning for nearly 100 years (e.g., Cashwell et al. 2001; Erlwanger 1973; Ormrod 2000; Stigler and Hiebert 1999). Teachers scribing a behaviorist belief tend to give students repetitive practice in the classroom in order to reinforce desired behaviors in the schooling process. Emphasizing routines and re-teaching is aligned with a repetitive practice approach to teaching. The emphasis of classroom management and organization is also linked to behavioral-oriented beliefs in teaching in high-poverty urban schools when there are low levels of social capital or of a sense of the school as a caring communities of learners (Muñoz and Vanderhaar 2006; Payne 1998); in other words, classroom management becomes the foundation for other important aspects of effective teaching such as student engagement and achievement (Marzano 2003; Saphier et al. 2008)).

In contrast, Chinese teachers' selections of (a) physically and emotionally safe environment, (b) feedback, and (c) teacher as a person as the most important factors that reflect Confucian values with respect to teaching and learning. Although Confucius did not overtly talk about maintaining a physically and emotionally safe environment for students, some of Confucius sayings do imply an emotionally safe environment in teaching. For instance, Confucius inculcated his disciples to care about students, to teach students with tireless zeal, to provide education for all students without discrimination (诲人不倦 and 有教无类, Confucian proverbs). If teachers treat students as Confucius described, students would be emotionally safe. In reference to feedback, Confucius contended that teachers must receive feedback from students' facial expression and from students' verbal expression in order to find the "right" moment to teach them (citation is the same as above). In modern classrooms, Chinese teachers have explored the way of in-class feedback appropriate to a large class (e.g., Shi 2010). Regarding the conceptualization about the teacher as a person, this element has been a continuing effort to improving teachers' morality in China' education for over 2000 years. Teachers are assumed to be role models for their students and to society in the Chinese context (e.g., Hu 2000).

Aside from the mean comparisons in each category on the ETQS, the Kruskal-Wallis analyses revealed 13 significant indicators. It was found that the majority of these indicators are aligned to constructivist beliefs regarding teaching and learning, which are present in both the USA and China's new curriculum state standards (e.g., NCTM 2000; CMOE 2001). Although both the USA and China's elementary school teachers learned these new beliefs in professional training programs, teachers have different preferences to select these beliefs as effective teaching indicators. Teachers may feel some constructivist beliefs are brand new and must pay attention to them. For instance, "high expectation" was present in both the USA and China's reform documents, but for US teachers, they may perceive this as an urgent agenda given that US students did not perform well when compared to Asian students in the large international studies (e.g., Fleischman et al. 2010; Mullis et al. 2003). Therefore, US teachers emphasized this indicator more than Chinese teachers did. Another example is "grouping." Teaching students in small groups is not new for US teachers, but it is new for Chinese teachers. There has been extensive research on cooperative learning in China since 2001 (Wang 2002). Now, Chinese students are expected to have small group activities rather than just sitting in the classroom and listening to a lecture. As a result, Chinese teachers emphasized this indicator more than US teachers. The last example is the indicator that considers student learning styles. This Western imported principle has been a slogan since 2001 China's curriculum reform (e.g., Chen 2007; Xiong and Li 2010; Liu and Zhao 2013). From 2000 to 2010, there were 13,476 peer-reviewed papers discussing student learning styles in China's educational discourse (Pan et al. 2012). The overwhelmingly emphasis on this principle may have influenced Chinese teachers' perceptions of learning styles when compared to their US counterparts.

One also needed to be aware of the most significant indicators among 13 significant indicators reported in the result section: grouping, higher-order skills, and re-teaching. Since re-teaching rarely happened in a China's class, it is reasonable that US teachers valued this indicator more than their Chinese counterparts. We recommend that researchers continue investigating the perceptions of grouping and higher-order skills to validate our finding. This may contribute to the "international borrowing" theory (Halls

1990). It was thought-provoking to see the two Western principles of grouping and higher-order skills were borrowed at different speed levels. That is, grouping have been emphasized more in China than the USA, and higher-order skills are still emphasized more by US teachers.

4.2 Perceptions of the effective teaching from different groups of the participants

Several findings in this study revealed teachers in different groups demonstrated different patterns of teaching effectiveness regardless nationality. Less effective teachers emphasized maintaining appropriate pacing of instruction in contrast to more effective teachers. This is reasonable since in any country more effective teachers always possess cognitive complexity and more sophisticated teaching skills than less effective teachers. More effective teachers may adjust their planned teaching lessons while assessing students' level of understanding of their teaching. For instance, if teachers feel students do not grasp the concepts well, they may spontaneously come up with new problems in class. On the other hand, less effective teachers may follow a planned lesson strictly without adequately assessing students' level of understanding. This finding implies that it is not adequate to categorize teachers as novice and experienced. Less effective teachers may have a number of years of teaching experience, but they still need training on teaching skills.

With regard to teaching experience, this study showed that elementary school teachers with 6–10 years of teaching emphasized more assessment skills than teachers with 20+years of teaching experience. This finding has two main implications. First, unlike other basic teaching skills, assessment skills are difficult for elementary school teachers to master. It takes up to 5 to 10 years to master these skills, depending on the professional learning context. Second, elementary school teachers with 20+years of teaching experience have mature on their assessment skills. As a result, researchers hypothesize that assessment skills may be an important factor that affect teacher effectiveness. New investigations on this issue may become part of future research both at the national and international levels.

Teachers from low and high socioeconomic schools demonstrated different priorities on four indicators when comparing total means. Teachers in low socioeconomic schools ranked "Considers student learning styles" and "Classroom management & organization" as more important items than teachers in high socioeconomic schools. Traditionally, research (e.g., Muñoz and Dossett 2001) has shown that socioeconomic status (SES) is a strong predictor of student achievement. Although unfortunate from an equity perspective, the higher students' SES is, the higher their test scores will be. Therefore, teachers in low socioeconomic schools emphasize "Classroom management and organization" since discipline issues might be more present in the context of poverty (Payne 1998). Teachers in high socioeconomic schools ranked "Excitement" as a more important item than teachers in low socioeconomic schools; this may imply that teachers in high socioeconomic schools implement multiple teaching styles in contrast to teachers in low socioeconomic schools. For example, seeking excitement does not always go to a routine such as lecture plus work-seat practice in mathematics class. It is also aligned with what Marzano (2007) claims of teaching as an art and as a scientific procedure.

When comparing different school levels, the total mean differences between Chinese high school teachers in the study of Meng et al. (2015) and Chinese elementary school teachers in this study were found comparable. The least means were all the same in the six categories. However, there were different perceptions when selecting the least important means in each dimension. In particular, in the second dimension, elementary school teachers selected "pacing" as the least important indicator, in contrast to "link instruction to real life" for high school teachers. In the third dimension, elementary school teachers perceived "uses a variety of questioning techniques" as the least important indicator, while high school teachers perceived "grouping" as the least important indicator.

4.3 Limitations of this study

There are several limitations in this study. First, the two samples from China and the USA did not match perfectly. For instance, The US data set was collected in one way (online survey) whereas the Chinese data set was collected in an entirely different manner (completing an in-person survey). It is also complicated to obtain matching samples in international studies regarding school locations, economic status of the schools, and school cultures. Second, we were not able to find enough male teachers in this study. As a result, the gender differences were only presented in China's sample. Third, the sample size in each country is also not big enough for more additional comparisons. Fourth, more qualitative studies might be needed to understand the cognitive complexity of the art and science of effective teaching. Further research may consider fixing these limitations while adding to the extant literature. This study is a research call for more international comparison studies that will help us fully understand the "black box" (Muñoz 2005) of what really constitutes effective teaching.

5 Final remarks

The comparison between the USA and China's elementary school teachers' perceptions regarding effective teaching indicated that students' engagement was the common highest priority and most important characterization of effective teaching in both countries. This is explained in part by the new rigorous standards that make student engagement a key prerequisite for the cognitive demands of higher-level thinking, creative problem-solving, and application of concepts to novel situations. Maximized student engagement can be critical for academic learning time. Engaged student time, a subset of allocated time, is the only time when students are really paying enough attention to actually learn from an instructional activity. This important finding of this study is supporting prior international comparative research by Teddlie and Liu (2008) that explored effective teaching and found that providing opportunities for student involvement is critical.

Differences were also found between the two countries teachers' perspective toward the conceptualization of effective teaching. The differences in educational context help in understanding the discrepancies between the two countries. In the US setting, particularly in large urban districts, there is a strong influence of behaviorist beliefs. In the Chinese setting, there is an established influence of moralistic beliefs rooted in

Confucianism. Regardless of the differences in educational context, the authors consider that the Eastern and Western educational systems can do better on the teaching-and-learning processes by continuously learning from each other and recognizing the value of their complementary way of approaching effective teaching. Just as skillful teachers reach out to their colleagues to improve their teaching practices, educational systems from around the world—like the USA and China—need to keep constantly growing in knowledge, skills, and dispositions related to teacher effectiveness. Educational systems can be learners too.

References

Black, P., & Dylan, W. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Cashwell, H. T., Skinner, H. C., & Smith, S. E. (2001). Increasing second-grade students' reports of peers' prosocial behaviors via direct instruction, group reinforcement, and progress feedback: a replication and extension. *Education and Treatment of Children, 24*(2), 161–175.

Chen, Y. (2007). Probe into the diversification of learning styles with correspondence to the new development orientation. *Curriculum, Teaching Material, and Method, 27*(3), 16–21.

China Ministry of Education. (CMOE). (2001). *Mathematics curriculum standards: grades 1 to 9*. Beijing, China: Beijing Normal University Press.

Connor, C. M., Jakobsons, L. J., Crowe, E. C., & Meadows, J. G. (2010). Instruction, student engagement, and reading skill growth in reading first classrooms. *The Elementary School Journal, 109*(3), 221–250.

Devine, D., Fahie, D., & McGillicuddy, D. (2013). What is 'good' teaching? Teacher beliefs and practices about their teaching. *Journal of Irish Educational Studies, 32*(1), 83–108.

Dillman, D. A., Smyth, J. D., & Melani Christian, L. (2009). *Internet, mail, and mixed-mode surveys: the tailored design method*. Hoboken, NJ: Wiley.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *The Journal of Mathematical Behavior, 1*(2), 7–26.

Fan, L., Wong, N., Cai, J., & Li, S. (2004). *How Chinese learn mathematics*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., & Shelley, B.E. (2010). *Highlights from PISA 2009: performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context (NCES 2011–004)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Grant, W. L., Stronge, H. J., & Xu, X. (2013). A cross-cultural comparative study of teacher effectiveness: analyses of award-winning teachers in the United States and China. *Journal of Educational Assessment Evaluation and Accountability, 25*, 251–276.

Grant, W. L., Stronge, H. J., Xu, X., Popp, P., Sun, Y., & Little, C. (2014). *West meets East: best practices from expert teachers in the U.S. and China*. Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (2007). Multiple sources of evidence: an analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measures: Issues and Practice, 26*(1), 19–27.

Halls, W. D. (1990). *Comparative education: contemporary issues and trends*. London: Jessica Kingsley Publishers.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin.

Hu, X. (2000). Reflection on teachers becoming a role model. *Educational Research, 9*, 2–5.

Kan, W. (2013). An empirical study on effective teaching–learning assessment system. *Education Research Monthly, 1*, 68–73.

Kirshner, D. (2002). Untangling teachers' diverse aspirations for student learning: a crossdisciplinary strategy for relating psychological theory to pedagogical practice. *Journal for Research in Mathematics Education, 33*(1), 46–58.

Liu, Z., & Luo, S. (2013). Characteristics of high school effective teaching in China. *Theory and Practice of Education, 33*(20), 43–46.

Liu, S., & Meng, L. (2009). Perceptions of teachers, students, and parents of the characteristics of good teachers: a cross-cultural comparison of China and the United States. *Educational Assessment, Evaluation and Accountability, 21*, 313–328.

Liu, S., & Zhao, D. (2013). Teacher evaluation in China: latest trends and future directions. *Educational Assessment, Evaluation and Accountability, 25*, 231–250.

Long, Y. (2010). High efficiency teaching strategies in history class. *Journal of the Chinese Society of Education, 12*, 74–75.

Ma, L. (1999). *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States* (p. 1999). Mahwah, N.J.: Lawrence Erlbaum Associates.

Marzano, R. J. (2003). *Classroom management that works: research-based strategies for every teacher.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007). *The art and science of teaching: a comprehensive framework for effective instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

Meng, L. Munos, M., & Wu, D. (2015). Teachers' perceptions of effective teaching: a theory-based exploratory study of teachers from China. *Educational Psychology.* doi:10.1080/01443410.2015.1008402.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2003). *TIMSS 2003 international mathematics report.* Boston: TIMSS & PIRLS International Study Center Lynch School of Education.

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., & Galia, J. (2008). *TIMSS 2007 international mathematics report: findings from IEA's trends in international mathematics and science study at the fourth and eighth grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muñoz, M. A. (2005). *Black box. Encyclopedia of evaluation.* Thousand Oaks, CA: Sage.

Muñoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: a longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education, 20*, 147–164.

Muñoz, M. A., & Dossett, D. (2001). Equity and excellence: the effect of school and sociodemographic variables on student achievement. *Journal of School Leadership, 11*, 120–139.

Muñoz, M. A., & Vanderhaar, J. E. (2006). Literacy-embedded character education in a large urban district: effects of the child development project on elementary school students and teachers. *Journal of Research in Character Education, 4*, 27–44.

Muñoz, M. A., Prather, J. R., & Stronge, J. H. (2011). Exploring teacher effectiveness using hierarchical linear models: student- and classroom-level predictors and cross-year stability in elementary school reading. *Planning and Changing, 42*(3/4), 241–273.

Muñoz, M. A., Scoskie, J. R., & French, D. L. (2013). Investigating the "black box" of effective teaching: the relationship between teachers' perception and student achievement in a large urban district. *Educational Assessment, Evaluation and Accountability, 25*(3), 205–230.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* Reston, VA: The National Council of Teachers of Mathematics, Inc.

Olejnik, S. F. (1984). Planning educational research: determining the necessary sample size. *Journal of Experimental Education, 53*(1), 40–48.

Ormrod, E. J. (2000). *Educational psychology: developing learners.* Upper Saddle River, NJ: Prentice Hall.

Pan, H., Luan, X., & Wang, Z. (2012). Review of learning styles. *Journal of Neijiang Normal University, 27*(7), 36–40.

Payne, R. K. (1998). *A framework for understanding poverty.* Highlands, TX: RFT Publishing.

Saphier, J., Haley-Speca, M. A., & Gower, R. (2008). *The skillful teacher: building your teaching skills.* Acton, MA: Research for Better Teaching.

Shi, Y. (2010). Reflection of the questions teachers asked in mathematics class. *Theory and Practice of Education, 20*, 53–54.

Stigler, W. J., & Hiebert, J. (1999). *Teaching gap: best ideas from the world's teachers for improving education in the classroom.* New York: Free Press.

Stronge, J. H. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Stronge, J. H. (2010). *Effective teachers = student achievement: what the research says.* Larchmont, NY: Eye On Education.

Stronge, J. H. (2013). *Stronge teacher evaluation system: a validation report.* New Jersey, USA: Stronge Associate Educational Consulting, LLC.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connections between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*(4), 339–355.

Teddlie, C., & Liu, S. (2008). Examining teacher effectiveness within differentially effective primary schools in People's Republic of China. *School Effectiveness and School Improvement, 19*, 387–407.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer.

Wang, T. (2002). The basic principle of cooperative learning. *Journal of Educational Research, 2*, 68–72.

Wang, Z. (2010). Exploring effective strategies of teaching high school mathematics. *Science Education, 16*(2), 35–37.

William, R.E. (2010). *Administrator and teacher perceptions of the qualities of effective teachers*. Unpublished dissertation. Virginia, USA: the College of William and Mary.

Xiong, Q., & Li, T. (2010). Exploring the relationships between teaching style and learning style. *Journal of Science and Education, 4*, 29–30.

Zhang, Z., & Deng, Y. (2010). A survey on effectiveness of high school chemistry teaching under the new curriculum reform. *Journal of Educational Development, 2*, 82–85.