Universiteit
Leiden
The Netherlands

Teachers' teaching and learning motivation in China

Zhang, X.

Citation
Zhang, X. (2021, July 6). *Teachers' teaching and learning motivation in China*. *ICLON PhD Dissertation Series*. Retrieved from https://hdl.handle.net/1887/3195071

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3195071

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle https://hdl.handle.net/1887/3195071 holds various files of this Leiden University dissertation.

Author: Zhang, X.
Title: Teachers' teaching and learning motivation in China
Issue Date: 2021-07-06

2

WHAT IS GOOD TEACHING: A COMPARISON OF THE PERCEPTIONS ON TEACHING QUALITY OF EXTERNAL SUPERVISORS AND STUDENTS

Abstract

New Basic Education (NBE) has been launched by university supervisors to encourage school teachers to improve their teaching. In NBE, assessment by university supervisors, who visit schools for their evaluation and feedback, is a crucial element. Besides supervisors' comments, as the consumer of NBE, students' voices should not be ignored. However, little is known about how university supervisors' assessments align with evaluations from primary school students. This study aims to fill this gap by exploring supervisors' and students' evaluations of 20 primary school teachers who participated in the USP. Their teaching practices were evaluated by 10 academic supervisors from NBE and 497 students from primary schools to explore the relationship between students' and supervisors' evaluations of teaching. In general, the results reported fairly low correlations between students' and supervisors' evaluations of teaching. It seems that students and supervisors applied different criteria and focused on different aspects of teaching. Students seemed to be more focused on learning climate, activating teaching, and instructional adaptation, whereas supervisors seemed to pay more attention to classroom management, instructional clarity, and strategies of instruction. Given both observations and student surveys have strengths and weaknesses, both methods should be seen as complementary ways to evaluate teaching.

This Chapter has been submitted for publication in an adapted form as Xin, Z., Saab, N., & Admiraal, W (under review). Students and supervisors' perceptions of teaching in primary education.

2.1 Introduction

In China, many large-scale education reforms have been launched as a result of the strong demand for teachers to learn how to improve their teaching to meet the challenges of a fast-changing society. One of these reforms is the New Basic Education (NBE), which can be understood as a form of university-school partnership – supervisors from universities visit teachers at schools to help them to reduce teacher-centred teaching in favour of student-centred approaches, and support them in becoming more inquiry-oriented and to engage students in solving practical problems, developing the ability of critical thinking (Li, 2020; Ye & Cheng, 2018). In teachers' continuous development, the partnership between school and university is a way to close the gap between theory and practice, and supervisors' assessment of teaching practices is a crucial element of the programme (Vrijnsen-de Corte, den Brok, Kamp, & Bergen, 2013).

Besides supervisors' measurement, as part of this collaboration, students' assessment of teaching quality should also be considered an important source of the programme, since that the ultimate goal of NBE reform is student achievement, and students are the consumers of their teachers' classroom teaching, their voices should be heard (Dockterman, 2017b).

Previous research reports that external observers and school students have access to different features of learning environments (Seidel & Shavelson, 2007). This indicates that the two sources of evaluation are important sources for studying learning environments. Some of these studies used quantitative data to explore the relationship between observer ratings based on classroom observation and student ratings measured in a survey (e.g., Dobbelaer, 2019a; Maulana & Helms-Lorenz, 2016). Such quantitative measures may help us to recognise the relationship between classroom observation and student survey, and the differences in specific aspects of instructional quality. However, it fails to tell us *why* students and observers have different perspectives on these specific aspects, and what are their respective evaluation criteria. In addition, most studies focus on students from high education, the research in primary education is quite limited. The present study focuses on comparing university supervisors and primary school students' evaluations of teaching, and uses mixed methods to not only explore the relationship between students' and external supervisors' perceptions of instructional quality, but also to interpret differences.

2.2 New Basic Education

In China, educational assessment is dominated by high-stakes examinations. In light of the importance of examination success, teachers training programmes organized by the government are designed to emphasize knowledge delivery, memory-driven learning, and teacher-centred approaches (Xin & Fred, 2014), which is always criticized for limiting students' deep learning and creativity (Yu, Chen, Levesque-Bristol, & Vansteenkiste, 2016). To counteract the test-oriented education, New Basic Education is designed by researchers from East China Normal Universities (ECNU). ECNU collaborative with partner schools to equip school teachers with new pedagogical knowledge to reduce teacher-centred teaching in favour of student-centred approaches under the facilitation of university supervisors. To ensure the successful implementation of new teaching strategies, supervisors from ECNU go to their partner schools regularly to discuss with teachers, deliver lectures, observer lessons, as well as provide professional recommendations (*see Introduction Table 1.1*). After years of development, more and more schools tend to seek professional support from university supervisors who are perceived as knowledge providers. NBE has become an increasingly influential project with the expansion of teachers' knowledge base (Bu & Han, 2019).

2.3 The role of supervisors in New Basic Education

Supervisors from the NBE have three different backgrounds: 1) theoretical researchers from ECNU, 2) teacher educators from local colleagues, and 3) part-time researchers from other universities (Li, 2020). They formed consulting teams to go to their partner schools regularly to assist teachers in schools to implement new teaching strategies by coaching and mentoring teaching practices. After classroom observation, a meeting will be organized and supervisors will give their comments on teaching based on the observation form, and teachers reconstruct their knowledge by being supported by trained supervisors.

Although many researchers have explored school administrator and principal evaluations of teaching (Supovitz et al., 2010; Yan, 2015), little is known about the evaluations of supervisors from universities. Moreover, unlike school administrators and principals, most supervisors are professors, researchers, or teacher educators

employed at universities, which might mean that they are more informed by international literature on teaching and learning that stresses the importance of student-centred teaching. However, the priority of administrators and principals is to ensure high student academic achievement in public examinations. The different interests may also lead to different criteria used in the evaluation of teaching. To fully understand the possible benefits of NBE and the role of supervisors in it, more insights are needed about the criteria supervisors use to evaluate teaching.

2.4 Evaluation of teaching quality in primary education

In the 1995 TIMSS video study, Klieme, Pauli, and Reusser (2009) develop a theoretical framework to elaborate teaching quality. Based on the framework, Fauth et al. (2014) present a model which can successfully be applied to evaluate teaching in primary schools. It consisted of three dimensions of teaching quality: 1) classroom management, 2) cognitive activation, and 3) supportive climate.

Classroom management is a well-known concept in educational research, it is operationalised how teachers deal with disciplinary problems and disruptions in the classroom. This classroom practice can be treated as preconditions for time on a task that is, in turn, significant for student achievement. Cognitive activation was related to the exploration of concepts, ideas, and prior knowledge. These classroom practices would develop students' cognitive engagement, in turn, lead to elaborated knowledge. Supportive climate refers to specific aspects of a positive teacher−student relationship and constructive teachers' feedback. It comprises teachers' warmth, encouragement, and constructive feedback. Classroom with a supportive climate that can fulfil students' needs and have positive effects on student outcomes. To enhance a positive climate, teachers should provide extra help when needed, respect students' questions and care about the students.

A growing global literature reported that claimed that these three dimensions reflect the key aspects of teaching, and can be replicated in ratings of students from primary schools, and are positively related to student academic achievement and subject-specific interest (Fauth et al., 2014).

2.5 Classroom observation and student survey to evaluate teaching

A considerable debate with regard to evaluating the teaching behaviour deals with selecting methods that are powerful enough to reflect 'real' teaching practice. Two common methods for measuring the teaching practice are registration methods (i.e., classroom observation) and methods based on perceptions, mostly student surveys (Hassan & Wium, 2014). Each method has its own strengths and weakness, below we will first discuss each method, and then go into the relationship itself between these two methods.

2.5.1 Classroom observation

The record from the observer is often considered as the most objective by many researchers (Dobbelaer, 2019b; Maulana & Helms-Lorenz, 2016). Observers can provide valuable information to teachers. However, it is not self-explanatory that classroom observation can always provide a valid evaluation. Several issues, such as the quality of external observers and the number of lessons per teachers that should be observed, need to be taken into account when using classroom observation. It can make classroom observation a costly and time-consuming method. In addition, the presence of observers can influence teachers' behaviour as well (Maulana & Helms-Lorenz, 2016).

2.5.2 Student survey

Compared to classroom observation, methods based on self-reports mostly use student perceptions to evaluate teaching can also function as a valuable source of feedback to teachers as they are the learners and spend the most time in the classroom (Dockterman, 2017a). In addition, perceptions from students are based on day-to-day experiences with the teacher during different lessons, not merely from a single or limited number of observations. The weakness of student survey is that several studies report that the evaluation has the potential to be influenced by teachers' personal factors which are unrelated to teaching quality, such as teacher popularity and teacher gender (Hassan & Wium, 2014; Wagner et al., 2013). However, there are many studies that confirmed the reliability and validity of student ratings (Aditomo & Koehler, 2020; Fauth et al., 2014).

2.5.3 The relationship between classroom observation and student survey

Studies that include both classroom observations and student surveys have shown low to moderate agreement between students' and external observers' evaluations of teaching (Dobbelaer, 2019b; Maulana & Helms-Lorenz, 2016). For example, in their study of ninth grade science classrooms, Lawrenz, Huffman, and Robey (2003) used a regression model to explore the relationship between supervisor and student evaluations of science classroom practice. The results indicated low correlations between observer observations and student evaluations. These low correlations are also supported by Maulana and Helms-Lorenz (2016). In that study, quantitative data was collected from 2,164 students of 108 teachers in the Netherlands. The results indicated a low agreement between trained supervisors and students in three aspects of teaching behaviour: learning climate, classroom management, and clarity of teaching. Some issues may have contributed to the low correlations that were found in the previous studies, such as different construction of measurements, different standards of external observers and students, or the fact that student questionnaire and classroom observation were not conducted at the same moment in time(Maulana & Helms-Lorenz, 2016).

2.6 This study

This study not only explores the relationship between students' and supervisors' evaluations of teaching, but also interprets the construct representation and potential differences measured by students and supervisors. We expect low correlations between perceptions from supervisor observations and student questionnaires because we anticipate that classroom observations by observers and student surveys of instructional behaviour are not simply different methodological approaches; rather, the two methods might be tapping into different representations of the meaning of 'real' instructional behaviour. Based on these considerations, the following research questions are addressed:

1. What is the relationship between supervisors' and students' evaluations of instructional quality?

2. What are the evaluation criteria used by supervisors and students?

2.7 Method

2.7.1 Participants

In this study, 12 primary schools that participated in NBE participated in this study. Many studies have indicated that the sense of self-efficacy is a significant predictor of teaching in the classroom (Gaertner & Brunner, 2018; Tschannen-Moran & Hoy, 2001). Accordingly, a total of 472 teachers completed a questionnaire on feelings of self-efficacy. They were numbered and divided into three groups based on their mean scores in self-efficacy: top 30% of teachers in the high-level group, then 40% of teachers in the medium-level group, and last 30% of teachers in the low-level group. Then we randomly selected 6 teachers from high-level-group, 8 teachers from medium-level-group, and 6 teachers from low-level-group to participate in the current study, with a total of 20 teachers. In addition, we invited 497 students of the 20 teachers to evaluate their teaching. These teachers were on average 34 years old (*SD*= 7.24) and taught different subjects. Sample statistics regarding age, gender, teaching experience, and other information are presented in Table 2.2. These students were on average 10.55 years old (*SD* =1.19). Class sizes range from 15 to 35 students per class, with a mean of 25 students (SD = 5 students).

Participation in the study was strictly voluntary and confidential for teachers, supervisors, and students. Upon recruitment, principals authorized the study within their schools, and teachers, supervisors, students, and their parents were asked to sign an informed consent regarding their collaboration in the study. Ethics approval for this study was granted by the institution the authors are from. Teachers completed a questionnaire in their offices. Students were given enough time to respond in classrooms. Supervisors were asked to finish their observation form in their offices after watching the video-tapes of the three lessons.

Table 2.2

Sample statistics

Teacher's number	Age	Gender	Subject	Teaching experience in years
366	30	Female	Math	7
23	30	Female	Math	3
428	30	Female	Math	6
65	37	Female	Math	14
16	35	Male	Math	12
63	44	Female	Math	21
96	46	Female	Math	4
208	30	Female	Math	5
353	45	Female	Chinese	22
114	49	Female	Chinese	26
283	30	Female	Chinese	7
226	30	Female	Chinese	7
42	29	Male	Chinese	7
228	23	Female	Chinese	1
315	30	Female	Chinese	6
46	42	Female	English	19
431	35	Female	English	12
108	36	Female	English	13
333	40	Female	Art	17
217	27	Female	Music	4

2.7.2 Procedure

Many studies have indicated that to get reliability information, there should be a limit of a total of three observations, and the interval between observations should be short to ensure stability in teaching quality (Hill et al., 2012; Maulana & Helms-Lorenz, 2016). Therefore, for each teacher, three lessons within three weeks were videotaped.

Based on teacher participation, we selected 10 supervisors to participate in this study. All supervisors had a Bachelor's or Master's degree in Education, rich teaching experience in primary education, and expertise in different subjects. Supervisors and teachers did not know each other. Teachers are assigned to different supervisors depend on the subjects they teach (see Table 2.3). The 10 supervisors rated 20 teachers' videos and provided comments according to the observation form. Then we invited students of the 20 teachers to evaluate their teaching. After three lessons, stu-

dents were asked to complete the questionnaire to indicate their evaluation of three lessons.

Table 2.3

The assignment of teachers to supervisors

Supervisor's number	Subject	Teacher's number
Supervisor 1	Math	366, 23,
Supervisor 2	Math	428, 65
Supervisor 3	Math	16, 63
Supervisor 4	Math	96, 208
Supervisor 5	Chinese	353, 114
Supervisor 6	Chinese	283, 226
Supervisor 7	Chinese	42, 228, 315
Supervisor 8	English	46, 431
Supervisor 9	English	108
Supervisor 10	Art and Music	333, 217

2.7.3 Measures

2.7.3.1 Student questionnaire

According to the framework created by Fauth et al. (2014), a paper-and-pencil student questionnaire was developed. It consisted of 21 items. All the items were adapted and reworked for application in primary school classrooms. We avoided negative formulations, inverted items, and complex expressions. Each item was scored on a 4-point Likert-type scale (1= strongly disagree; 4= strongly agree). The 21 items were subject to an exploratory principal component factor analysis with direct oblimin to determine underlying factors. The final analysis consisted of two components of 19 items, which explained 41.9% and 11.5% of the variance in scores, respectively. Two items were deleted because of low factor loadings and no cross-loadings (>.40) were found. The first component included cognitive activation and supportive climate, which was labelled to "classroom teaching", indicating exploration of students' prior knowledge and building a supportive learning climate. The second component was labelled "classroom management" with items on classroom rules and procedures dealing with disruptions. The Cronbach's alphas of classroom teaching and classroom management were 0.91 and 0.89, respectively. All the items of the

student questionnaire are shown in Appendix 1.

2.7.3.2 Classroom observation

To better capture how supervisors evaluate teaching quality, we decided to use the data from observation form that is used in the programme. The supervisors have been trained for many years to use this observation form to evaluate teaching quality. We assumed that using this form could help supervisors to provide us more insights into the evaluation of teaching quality. In addition, this observation form is currently used in more than 200 schools. Using this form will help us to better understand what is happening in the classroom, and how supervisors modelling school teaching in practices.

Similar to the student questionnaire, the data from observation forms is also divided into two main categories: classroom management, and classroom teaching. For classroom teaching, five subscales have been distinguished:1) Safe and stimulating climate, 2) Clear instruction, 3) Activating teaching, 4) Teaching learning strategies, 5) Adaptation of teaching. Each supervisor rated each item on a 5-point scale (1= insufficient; 5= good) and completed an open-ended question for each subcategory where they could give a more detailed account of their opinions. The supervisor rating form is included in Appendix 2.

Although the observation instrument provides more categories than student questionnaire, they both include classroom management and classroom teaching. It becomes possible to compare these conceptions theoretically.

2.7.4 Reliability and validity

The reliability and construct validity of the structure of the student questionnaire were tested previously by research conducted by Fauth et al. (2014). They gathered questionnaire data from 1556 primary school students (third grade), the analyses show that student ratings can be treated as useful measures of teaching quality in primary school.

For the supervisor rating form, which is used as a common tool to evaluate teaching for more than 200 schools in 14 Chinese cities, showing good reliability, validity,

and intercultural equivalence (Bu & Han, 2019; Li, 2020).

2.7.5 Videotape procedures

To record teaching in the classroom, teachers' lessons were videotaped following the standardized procedures used in the TIMSS Video Study (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999). Videographers shot each lesson using a single camera in the classroom, capturing teaching behaviours during the lesson. The camera was turned on at the beginning of the class and turned off when the lesson was finished, totally about 35 minutes. Teachers were told that the purpose of the study was to videotape typical lessons, and that they were asked not to make any special preparations for the lesson.

2.7.6 Training of students

Students were instructed about the questionnaire in a session of 30 minutes by the first author. The training involved explanations of the student questionnaire and how to evaluate teaching practices using associated scoring rules. After that, students were given enough time to complete the questionnaire in class. The first author stayed nearby and gladly answered any questions.

In this study, we used the supervisors' observation form as our measurement to evaluate teaching. Since supervisors have used the observation form for many years, it seems that the training and computation of reliability measures using the data from the training sessions were inappropriate.

2.7.7 Analysis

Comparing average sum scores is very common in educational research. We first created four variables by calculating the mean scores for classroom management and classroom teaching according to students' and supervisors' evaluations. They were labelled as: students' evaluations of classroom management; students' evaluations of classroom teaching; supervisors' evaluations of classroom management; and supervisors' evaluations of classroom teaching. To explore the relationship between students' and supervisors' evaluations of teaching, a correlation analysis was conducted to measure the strength of association between these variables. A sample t-test was

carried out to determine whether the mean scores significantly differed.

Secondly, to find to further examine the relationship between the evaluation of classroom management and classroom teaching by supervisors and students, we created two cross tables, one for classroom management (Table 2.5), and one for classroom teaching (Table 2.6). In these tables, the 20 teachers were clustered into nine groups, referring to high, medium or low evaluations from either students or supervisors according to their mean scores on each variable (30% of teachers were divided into the high-level group, 40% of teachers were divided into the medium-level group, and 30% of teachers were divided into the low-level group). Finally, 6 teachers were assigned to high level-group, 8 teachers were assigned to medium-level-group, and 6 teachers were assigned to low-level-group.

Thirdly, we selected the teachers with different evaluation scores from students and supervisors and created a table to summarize supervisors' qualitative comments for these teaching according to the observation form. In the table, we provided a comprehensive view by comparing the strong points and weak points of these teaching according to the supervisors to explore the differences between students' and supervisors' evaluations of teaching.

2.8 Results

Table 2.4 shows the results of Pearson correlation between the evaluations of students and supervisors. It indicates that there was a significant positive relation between students' evaluations of classroom management and classroom teaching ($r = 0.77$, $p < .001$), and between supervisors' evaluations of classroom management and classroom teaching ($r = 0.76$, $p < .001$). No significant relationships were found between students' and supervisors' evaluations of the same aspect of teaching: for classroom management ($r = 0.13$, $p = 0.594$). And for classroom teaching ($r = -0.03$, $p = 0.900$).

Table 2.4

The correlations of students' and supervisors' evaluations of teaching (N=20)

	Mean	*SD*	1	2	3	4
1: Students' evaluations of classroom management	3.73	0.63	-			
2: Students' evaluations of classroom teaching	4.20	0.38	0.77**	-		
3: Supervisors' evaluations of classroom management	3.10	0.64	-0.02	-0.03	-	
4: Supervisors' evaluations of classroom teaching	2.92	0.49	0.13	0.05	0.76**	-

A paired t-test was conducted to compare students' and supervisors' evaluations. A significant difference was found for both teaching aspects. Students rated both classroom management and classroom teaching significantly more positive than supervisors (classroom management, $t(19) = 3.36$, $p = 0.003$; and classroom teaching, $t(19) = 8.77$, $p < 0.001$).

2.8.1 Students' and supervisors' evaluations of classroom management

As mentioned above, we have created a cross table to cluster the nine groups of 20 teachers, referring to high, medium or low evaluations from either students or supervisors. Some teachers were rated high on classroom management, whereas supervisors rated them at a medium level. According to supervisors' comments (see Appendix 3), these teachers showed some basic management skills. However, the supervisors mentioned these teachers showed too many controlling behaviours in class. Some supervisors also mentioned the poor management skills of these teachers.

In addition, some teachers were rated at the medium level by their students, but at the low level based on their supervisors. For the teachers with a low level, supervisors indicated that these teachers failed to ensure the orderly progression of the lesson.

Finally, our results also indicated that some teachers were rated at a low level by students, but at the medium level by supervisors. Supervisors reported these teachers showed some basic management skills. But the supervisors noticed that the progres-

sion of a lesson was too fast to allow students to follow teachers' ideas.

Table 2.5

Teachers who were rated at a high level, medium level and low level by students and observers in terms of classroom management

	Supervisor evaluation		
	High	Medium	Low
Student evaluation			
High	23	16, 217, 366, 353	
Medium	228, 428	114, 63, 333, 283	431, 108, 226, 96
Low		315, 42, 65, 208	46

As shown in Table 2.5, some teachers were rated as medium level by both students and supervisors; these teachers were reported as providing some primary management skills. For the teacher who was rated as low level by both students and supervisors (teacher 46), supervisors reported she showed too many controlling behaviours and failed to provide an orderly progression in her lesson.

2.8.2 Student and supervisor perceptions of classroom teaching

As shown in Table 2.6, some teachers were rated high on classroom teaching by their students, whereas supervisors rated at the medium level. Almost all supervisors mentioned that these teachers provided a good learning climate and designed good interactive activities. However, supervisors reported that these teachers failed to involve all students in their lessons, and did not give clear instructions and explanations. In addition, supervisors also suggested these teachers needed to provide more timely help for struggling learners and to use more strategies to motivate students for learning.

Three teachers were rated at the medium level by their students, but at high-level based on supervisors' evaluations. Generally, supervisors reported these teachers were doing very well in many aspects of teaching except creating a relaxed learning atmosphere. We also noticed that three teachers were rated as low level by their students, but at medium level by supervisors. These teachers were considered by their supervisors to show basic teaching skills but failed to build a good learning climate and to implement effective interactive learning activities.

We also noticed that one teacher (teacher 16) was rated as high level, five teachers (teachers 46, 428, 63, 283, 114) were rated as medium level, and two teachers (teachers 109, 431) were rated as low level by both students and supervisors. For the high-level teachers, supervisors reported these teachers showed excellent teaching skills. For the medium level teachers, they mentioned that, although they showed some basic teaching skills and instruction was clear, they still needed to pay more attention to include all students, not just those who study well. For the low-level teachers, supervisors reported that they showed poor teaching skills in the class and their teaching skills needed to be improved (see Appendix 4).

Table 2.6

Teachers who were rated at a high level, medium level and low level by students and observers in terms of classroom teaching

	Supervisor evaluation		
	High	Medium	Low
Student evaluation			
High	16	217, 333, 366,353	
Medium	23, 65, 228	46, 428, 63, 283,114,	226, 96
Low		42, 208, 315	108, 431

2.9 Discussion

In this study, we investigated teachers' classroom management and classroom teaching as evaluated by supervisors and students. Two major findings emerged. First, our results reported indicated students and supervisors might use different criteria to evaluate teaching. In addition, there is a fairly low correlation between students' and supervisors' evaluations of teaching.

2.9.1 Students' and supervisors' evaluations of teaching

First, our findings suggested that both students and supervisors have different views on teaching. Basically, students evaluated their teachers more positively compared to supervisors. In contrast, supervisors hold a more comprehensive, and a higher standard to evaluate teaching compared to students. Specifically, it seems that students and supervisors have different foci in evaluating teaching. Compared to the

aspects of learning climate and activating teaching, supervisors seemed to have more-favourable judgments regarding whether the teacher managed the classroom effectively, used teaching language clearly, and applied teaching strategies appropriately. Especially with respect to teaching strategies, almost all the supervisors mentioned that they paid more attention to check whether the teacher uses appropriate teaching strategies to foster students' ability of critical thinking and break down complicated problems. For supervisors, it seems that using effective teaching strategies is a key factor that determines the quality of teaching. Compared to supervisors, students seemed to be more focused on learning climate, activating teaching, and adapting of teaching. Students rated a teacher highly when the teacher designed a lot of interactions, constantly posed questions, timely praised students or adapted the assignments to relevant differences between students. We even noted that some teachers who failed to show clear instructions and explanations or apply teaching strategies appropriately, were still rated highly by students just as they created a good learning climate. Such findings led us to conclude that students' evaluations of behaviour are not to be interpreted without caution. The reason why students and supervisors have different focuses on the aspects of teaching might be attributable to different perspectives. For supervisors, as the teacher educator, they were trained to use the scoring rules to evaluate teaching, and wanted teachers to apply constructivist teaching approaches in their teaching. This makes them focus more on complex teaching behaviour (i.e., to involve all the students and to apply appropriate teaching strategies) than on some relatively simple teaching behaviour (i.e., to build a relaxed atmosphere), and hold more rigorous quality criteria for teaching. For students, their evaluation is a subjective evaluation based on learning experiences with teachers (Maulana & Helms-Lorenz, 2016). Compared to the complex teaching behaviour, some relatively simple and directive behaviour (i.e., to create a safe and relaxed climate and to provide clear explanations of an assignment) might be easier to recognise and attract their attention. Additionally, one of the purposes of NBE is to support learner-centred teaching (Yan, 2015). Therefore, we expected supervisors would encourage teachers to create a comfortable learning climate and give students more autonomy in the classroom. However, our findings suggest that supervisors paid less attention to the aspect of learning climate and were more focused on classroom management. Some teachers were even rated at a low level by supervisors as they mentioned these teachers provided too much autonomy to students. This finding contradicts some research in other countries that claimed that supervisors focus more

on a supportive learning climate than students (Maulana & Helms-Lorenz, 2016). A more in-depth study is needed because the reasons for these differences remain inconclusive. One possible explanation could be that supervisors may be afraid that creating a supportive learning climate would make the classroom noisy and busy, which affect students' academic performance negatively. As mentioned above, Chinese education is still dominated by high-stakes examinations .

2.9.2 The correlations between the students' and supervisors' evaluations of teaching

Our results indicated that there are fairly low correlations between the students' and supervisors' evaluation of teaching. Our results are compatible with the findings of Maulana and Helms-Lorenz (2016), who also reported a low agreement between students' and external observers' evaluations of teaching. As mentioned above, students and supervisors have different perspectives, which may make students and supervisors have different foci in evaluating teaching. Another possible reason is that compared to the student questionnaire, the observation instrument is more holistic regarding its wording and formulation, and provide additional items that are not present in the questionnaires for students since it is targeted at experienced supervisors. In contrast, the items of the student questionnaire are more specific and formulated to student perceptions in primary education. These differences in operationalization at the item level may also contribute to the low correlations. The low correlations mean that using either evaluation of students or ratings by supervisors may lead to a one-sided and incomplete view of teaching. To best determine the quality of teaching in a professional learning programme, it may be worth including these two measures in the evaluation system as they each provide a different view on teaching.

Finally, although they have different judgments regarding teaching, it seems that both supervisors and students care about whether the teacher involves all students in the classroom. Supervisors reported that most teachers have the ability to adapt their teaching to the differences between students. However, in practice, supervisors indicated that it seemed that some teachers only invited good learners to answer their questions and participate in learning activities, ignoring struggling learners.

2.9.3 Implications

Our findings can have implications for school leaders and policymakers to evaluate teaching.

First, our study suggests that primary school students rated their teaching higher than universities supervisors, and more-favourable judgments regarding learning climate, activating teaching and adaptation of teaching, however, may be less able to rate some specific aspects of teaching than supervisors (i.e., clarity of instruction and teaching strategies). Therefore, if principals want to use students' perceptions to evaluate teaching, they should be aware that the weakness of using evaluations of students. For the teachers who are rated highly by students, it might be necessary to invite supervisors to further rate their performance in the classroom, especially in the aspects which students are less focused.

Second, for the supervisors, compared to students' evaluation, their comments are more critical, and more able to rate whether the classroom is managed effectively, and teaching strategies are applied reasonably. It seems that supervisors' perceptions could be used to help teachers to develop their "senior skills" of teaching quality. However, considering using observation measure is costly and labour-intensive, most schools still use students' perceptions to evaluate teaching, supervisors' comments could be used as a supplement to assess to what extent teaching meets teaching standards in the specific aspects of teaching.

Finally, the results of the evaluation of teaching will differ substantially depending on which of the two rater-groups is used. Given both observations and student surveys have strengths and weaknesses, it might be important for principals to select different measures depending on different purposes. Summative and formative purposes are the most cited purposes of teacher evaluation in the literature (Stronge, 2006). When principals want to use summative evaluation of teaching, they might use student evaluations of teaching. When principals want to implement collegial consultation and peer feedback/assessment to improve teaching, it might be better to use supervisor comments. As a formative assessment, the supervisors' evaluation embeds assessment processes throughout the teaching process to constantly improve teaching. Teachers and principals can use these to find out how the teacher is teaching, as well as what they need to do next to move their teaching forward.

2.9.4 Limitations

Generally, there are three limitations that should be carefully considered when interpreting our conclusions.

First, our study only used quantitative methodologies to explore student evaluation of teaching. Such quantitative measures help us to recognise scores in the specific aspects of teaching. However, we still do not know why students gave such scores. We can only speculate about the reasons based on supervisor descriptions. It would have been informative to include student interviews into our study. Such information may provide us with a deeper understanding of students' evaluations of teaching. We advise future studies use both quantitative and qualitative methodologies to explore how students and supervisors perceive teaching, and compare similarities and differences in the specific aspects of teaching.

Second, in order to determine supervisors' evaluations of teaching quality, the selection of supervisors for the sample was very important. In this study, supervisors were selected carefully to arrive at a diverse sample. However, our samples relied on supervisors who volunteered to participate in this study. This selection might have influenced the findings.

Third, although stratified random sampling was used to select teachers, our sample was relatively small (N= 20 teachers), which might influence the generalization of findings.

2.9.5 Concluding remarks

In this study, the evaluations of the teaching of supervisors and students were compared, and the results showed significant differences. The differences provide us with the opportunity for a better understanding of how students and supervisors perceive teaching. Such knowledge would be beneficial for classroom researchers to determine if a particular method of data gathering would generally be useful for measuring teaching behaviour. It also indicates that the reality of the teaching situation is not based on one 'truth'; each source brings different perspectives on teaching. Using only evaluations of either supervisors or students may lead to a one-sided and incomplete view. Therefore, we recommend teachers, policymakers, and school

leaders to focus more on the selection of multiple measures to evaluate teaching behaviour and choose the appropriate method according to different purposes.