UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Southampton Education School

THE EFFECTIVENESS OF MATHEMATICS TEACHING: A CROSS-NATIONAL INVESTIGATION IN PRIMARY SCHOOLS IN ENGLAND AND CHINA

by

Zhenzhen Miao

Thesis for the degree of Doctor of Philosophy

August 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Southampton Education School

Doctor of Philosophy

The Effectiveness of Mathematics Teaching: A Cross-national Investigation in Primary Schools in England and China

by Zhenzhen Miao

This study focuses on the effectiveness of mathematics teaching to children aged 9-10 years, applies a mixture of six methods to classroom-level data collected in England and China, correlates observable teacher behaviours with pupil mathematics performance and collects multiple perceptions that indirectly connect with the differences of teaching and learning cross-nationally. It has been found in the study that 9- to 10-year-olds ($n = 343$) from China outscored their English peers ($n = 236$) at the same age by over 20 per cent in each of two mathematics tests derived from TIMSS 2003. Structured analysis of lesson videos has revealed that Chinese mathematics teachers scored much higher than their English colleagues on an internationally validated observation instrument which focused on the *quality* of six dimensions of teacher behaviours. Furthermore, the *quantity* of teacher behaviours was also measured and the subsequent correlational analysis on pooled data indicated a positive effect of whole-class interactive teaching ($r = 0.97$, $p < 0.01$) and pupil time on task ($r = 0.95$, $p < 0.01$) and a negative impact of whole-class lecture ($r = -0.91$, $p < 0.01$), individual/group work ($r = -0.81$, $p < 0.05$) and classroom management ($r = -0.77$, $p < 0.05$) on pupils' mathematics performance cross-nationally. Qualitative findings are connected with quantitative results to explain how teachers think, how this relates to the way they teach and how the differences of teaching result in the performance gap cross-nationally. The study replicated previous TER findings from the West across two geographically and culturally different countries, suggested possible directions for

future enquiries, and recommended potential ways for practice and policy innovations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DECLARATION OF AUTHORSHIP

I, Zhenzhen Miao, declare that the thesis entitled "*The Effectiveness of Mathematics Teaching: A Cross-national Investigation in Primary Schools in England and China*" is the result of my own work and has been generated by me as the result of my own original research. I confirm that:

- this work was done wholly while in candidature for a research degree at this University;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Part of this work has been published as:

Miao, Z., Reynolds, D., Harris, A. & Jones, M. S. (2015) Comparing performance: a cross-national investigation into the teaching of mathematics in primary classrooms in England and China. *Asia Pacific Journal of Education*. DOI: 10.1080/02188791.2015.1056593

Miao, Z. & Reynolds, D. (2015). Uncovering classroom-level factors for the performance gap in mathematics between England and China: a synthesis of results and findings from an international teacher effectiveness study. In, *The Annual Conference of British Educational Research Association (BERA)*, Belfast, GB, 15 - 17 September 2015.

Miao, Z. & Reynolds, D. (2014). The effectiveness of mathematics teaching: a cross-national investigation in primary schools in England and China. In, *The Annual Conference of British Educational Research Association (BERA)*, London, GB, 23 - 25 September 2014. (press released)

Signed: __________________*Z. Miao*・苗偵楨 **Date**: _______________

ACKNOWLEDGEMENT

Coming to the UK with complete disappointment at Chinese educational system, I now end up with finding China has got effective maths teachers, which seems so dramatic and unbelievable. This PhD journey is first of all a self-healing journey – a journey that unlocked almost every question I held regarding education back in Nanjing, a journey that transformed every bias I held regarding Chinese schools and teachers and English schools and teachers into scientific results, a journey that put every such bias back into balanced views through international dialogues between insiders and outsiders. It happens to be more than a personal journey, as it coincides with many individuals and groups' interests at a time when the world is so eager to know why China has taught children maths so well. Without the help from many people, this journey would not be possible. A few hundred words are not enough to fully express my gratitude to them. The following words could hardly express 1% of my gratefulness.

I wholeheartedly appreciate my Supervisor, Professor David Reynolds, a leading academic and most importantly a wise person. I have learnt so much from him both academically and philosophically. It is very common for PhD students to feel completely lost at many points. Professor Reynolds is like the light tower who keeps telling me, "You're heading in the right direction. Stay confident. Don't give up. Keep moving forward." Whenever a deep sense of blindness and self-doubt started to drown every bit of my courage and morale, he was the wise man who always saw the bigger picture and said that it was OK and no worries.

Thanks to his supervision, I have grown into a more knowledgeable, more independent and wiser person ready for future challenges in academia and beyond.

I also want to express huge gratitude to the University of Southampton and Southampton Education School for accepting my research proposal, offering me the Vice-Chancellor's Scholarship and Southampton Education School Award and providing me the supportive environment to design and carry out the study. Many thanks also go to the School's Maths and Science Education Centre for providing financial support to cover my travel expenses between Southampton and Nanjing. I am very grateful to the School's Digital Video Studio for offering great support in providing the filming equipment for the study.

There are a number of staff members in Southampton Education School that I own massive thanks too. I sincerely appreciate Professor Lianghuo Fan for recommending two professors in Nanjing who played crucial roles in seeking participants in the city. I would like to thank wholeheartedly Dr Julie-Ann Edwards for her earlier guidance during the MA period, her genuine support during the PhD scholarship application process and her thought-provoking supervision during the first six months of my PhD study. I also owe a big thank you to Professor Melanie Nind for her invaluable advice at times and for recommending Professor David Reynolds as my supervisor.

Enormous thanks go to pupils and teachers who spared their invaluable time to take part in the EMT project, school leaders who believe in the value of the project and have the faith that education makes a difference, Professors Lianhua Ning and Wenbin Xu at Nanjing Normal University who kindly helped contact local schools, and two Masters students at Nanjing Normal University, Ms Yunyun Du and Ms Lijun Wu, who generously helped with data collection in Nanjing schools, and my relatives in Nanjing, Junhong Hua and Yan Wang, for their effort in helping with the delivery and collection of test papers.

Last but not least, I would like to thank my parents, Daozhong Miao and Guiling Zhao, for their love, high expectations and support, thank my sister, Yuanyuan Miao, for sharing the same passion in education, thank my husband, Wei Wang, for his four years' hard work in this country to support the three of us and for listening to me and encouraging me every time I felt lost, thank my son, Yuchao Wang, for being so patient and supportive – he is my lifelong Supervisor who, at the age of 13 and upon my first academic interview, has wisely suggested me,

" Be yourself."

Finally, I would like to thank my family home and abroad as a whole for their love that transcends oceans to reach me here and now.

I dedicate my thesis to

Every Child

in this world

past, present & future

for they are the ultimate teachers who teach

teachers how to teach them best

All grown-ups were once children... but only few of them remember it.

- Antoine de Saint-Exupéry, The Little Prince

ACRONYMS & ABBREVIATIONS

CHC	*Confucian-heritage-cultural*
CN	*China or Chinese*
CN-FG	*Chinese focus group*
CPD	*continuing professional development*
DfE	*Department for Education (UK)*
DfEE	*Department for Education and Employment (UK)*
EER	*educational effectiveness research*
EMT	*effectiveness of mathematics teaching*
EN	*England or English*
EN-FG	*English focus group*
ERGO	*Ethics and Research Governance Online*
ETN	*the Effective Teachers of Numeracy project*
ETS	*the Educational Testing Service*
FIAC	*the Flanders Interaction Analysis Categories*
FIMS	*the First International Mathematics Study*
FSM	*free school meals*
GDP	*gross domestic product*
GVA	*gross value added*
IAEP	*the International Assessment of Educational Progress*
ICCAMS	*the Increasing Competence and Confidence in Algebra and Multiplicative Structures project*

IEA	*the International Association for the Evaluation of Educational Achievement*
ISCED	*International Standard Classification of Education*
ISERP	*the International School Effectiveness Research Project*
ISTOF	*the International System for Teacher Observation and Feedback*
itvQ	*interview question*
IWB	*interactive whiteboard*
KS2	*Key Stage 2*
LPS	*the Learner's Perspective Study*
M1	*Method 1*
M2	*Method 2*
M3	*Method 3*
M4	*Method 4*
M5	*Method 5*
M6	*Method 6*
MEPP	*the Mathematics Enhancement Project Primary*
MET	*the Measure of Effective Teaching project*
MM1	*Mixed Methods – Part 1*
MM2	*Mixed Methods – Part 2*
MMR	*mixed-method research*
NAEP	*the National Assessment of Educational Progress*

OECD	*the Organisation for Economic Co-operation and Development*
Ofsted	*the Office for Standards in Education*
ORACLE	*Observational Research and Classroom Learning Evaluation*
OTL	*opportunity to learn*
PACE	*the Primary, Assessment, Curriculum and Experience project*
PC	*personal computer*
PE	*Physical Education*
PISA	*the Programme of International Student Assessment*
PQ	*pupil questionnaire*
PRC	*People's Republic of China*
QUAL	*qualitative*
QUAN	*quantitative*
RQ1	*Research question 1*
RQ2	*Research question 2*
RQ3	*Research question 3*
SATs	*Standard Assessment Tests*
SD	*standard deviation*
SE	*school effectiveness*
SEN	*special educational needs*
SER	*school effectiveness research*

SES	*socio-economic status*
SI	*school improvement*
SIMS	*the Second International Mathematics Study*
SPSS	*statistical package for the social sciences*
STEM	*science, technology, engineering & mathematics*
TA	*teacher assistant*
TALIS	*the Teaching and Learning International Survey*
TER	*teacher effectiveness research*
TIMSS	*the Third International Mathematics and Science Study*
TIMSS	*Trends in International Mathematics and Science*
TQ	*the teacher questionnaire*
UK	*the United Kingdom*
UNESCO	*the United Nations Educational, Scientific and Cultural Organisation*
US	*the United States of America*
WB	*whiteboard*

CHAPTER 1 - INTRODUCTION

Chapter overview

- *Chapter introduction*
- Importance of mathematics
- Research rationale
- Research contexts
- Research purposes
- Research questions
- *Structure of the thesis*

1.1 Chapter introduction

In this increasingly digitalised and globalised world, the quality of education in mathematics and other STEM subjects predicts the competitiveness of a country's future workforce. International comparative studies on achievement have repeatedly identified the disparity in mathematics performance between East Asia and the West (Mullis et al., 1997; Mullis et al., 2008; Mullis et al., 2012; Mullis et al., 2004; Reynolds & Farrell, 1996) and between China and England (Lapointe et al., 1992; Lapointe et al., 1989; OECD, 2010, 2013). Evidence from empirical studies has consistently proved that schools make a difference and teachers have a much larger impact upon pupils' achievement than do schools (Hill & Rowe, 1995; Reynolds & Teddlie, 2000; Willms, 2000). It has been argued that it was necessary to get into the classroom and see what actually happened if one wanted to improve learning (Hiebert et al., 2003). This study thus set out to (1) investigate teaching factors that work across England and China and make a difference in pupils' mathematics learning outcomes, (2) understand how different roles in and beyond classrooms view mathematics teaching across two countries and (3) interpret the interconnections between hard evaluations of teaching and learning and soft perspectives of practitioners and the researcher.

In this introductory chapter, section 1.2 reviews the importance of mathematics, section 1.3 explains the rationale of the study, session 1.4 introduces the research contexts, research purposes are illustrated in section 1.5, section 1.6 presents the research questions (RQs), and finally section 1.7 shows the structure of the thesis.

1.2 Importance of mathematics

Mathematics, as one of the most important bodies of knowledge, has developed across cultures and over millennia. It is the oldest school of human thought, alongside philosophy (Krantz, 2010). The mathematical historian, Carl Boyer, affirmed in his work, “that the beginnings of mathematics are older than the oldest civilisations is clear” (Boyer, 1968, p. 7). As far as archaeologists have discovered, the formal application of mathematics can be traced back to the era of Old Egypt and Old Babylonia around 3200 BC (Nuffield Foundation, 1994, Ch.2). Since the era when Arabians drew together the two lines of mathematics – the East (ancient India) and the West (ancient Greece) – through translation and introduced it to Europe, mathematics has gradually become an international body of knowledge (Boyer, 1968; Hodgkin, 2005).

Throughout history, mathematics has come from humans’ observation of the world and guided human beings in tackling real-world problems, providing powerful tools to the development of all sciences. In today’s world, mathematics remains a crucial part of almost every branch of science, including physics, chemistry, biology, architecture and electronic and computer science. For its importance in the past, present and future, mathematics is regarded, in schools across the world, as a core subject that provides essential foundations for other school subjects. The importance of mathematics makes it important that children be well educated in the subject. Such an importance also results in the increase of the number of educational studies focussing on the quality or process of mathematics education over the past 50 years. The following section will explain why a new study like this is needed and why it was carried out.

1.3 Research rationale

International comparative surveys, conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD), have been regularly evaluating educational outcomes of core subjects, such as mathematics and science, cross-nationally. Although the target ages of pupils those international surveys choose to focus on are different from one another, they all include mathematics as an assessed subject. The results of international assessments draw much attention and raise questions as to how educational effectiveness differs amongst nations and cultures (Hiebert & Stigler, 2004). The number of international surveys on learning outcomes is much greater than that on teaching practices which have been heavily researched on a within-country basis.

Empirical evidence from teacher effectiveness research (TER) consistently reveals positive correlations of certain teacher behaviours with pupil learning outcomes/gains mainly in mathematics and other core subjects (Muijs & Reynolds, 2011). However, over the past four decades, such evidence has been limited to national levels. In this increasingly globalised world, there is a lack of international attempts to reshape educational research agendas, evaluate which teaching variables "travel" across nations and which do not, "tap the full range of variation", generate "more complex, sensitive and multi-layered explanations" of differences in learning outcomes, and subsequently inform classroom innovations, as for the case of school effectiveness research (SER) (see Reynolds et al., 2002b, pp. 5-6).

This study, therefore, seeks to fill the research gap by collecting and analysing data across countries to find up-to-date evidence on the effectiveness of mathematics teaching (EMT) that could travel confidently across countries as mathematics itself did, does and will undoubtedly continue to do.

Applying a combination of convenience and stratified sampling strategies, the study collects data in two countries – England and China – and involves maths teachers and their pupils aged 9 to 10 (i.e. Year 5 in England and Grade 4 in China). In each country, a city from the medium-to-high level stratum was chosen, and within each city (Southampton, England and Nanjing, China), schools were sampled from the average stratum. Whilst detailed sampling techniques, procedures and rationales are illustrated in Chapter 3, the following section draws a picture of the research contexts.

1.4 Research contexts

This section intends to (1) build comprehensive profiles for the participating nations, cities and schools, (2) explain the comparability of the two nations and two cities and (3) offer an overview of education in both countries.

1.4.1 England and China

As follows, a brief introduction is given to England and China, in terms of geography, population, culture, politics, education and economy. Comparability of the two countries is explained as well.

1.4.1.1 Geography

England is the largest of the four countries in the United Kingdom of Great Britain and Northern Ireland (the UK or Britain). The area of England is 130, 281 km^2 – more than a half of the total area of the UK (Office for National Statistics, 2013b). It is bordered by Scotland to the north and Wales to the west and surrounded by the Irish Sea to the northwest, the North Sea to the east, and the English Channel to the south. Divided by the 33-km English Channel from France, England is the nearest country of the UK to mainland Europe.

The People's Republic of China (generally called China) is located in East Asia, between latitudes 18º and 54ºN, and longitudes 73º and 135º E. With Beijing as its capital, China has a total area of 9.6 million km^2, which makes it the world's second largest country by land and the third or fourth by total area. It shares land borders with 14 countries and maritime borders with 8 countries (PRC Gov., 2013).

1.4.1.2 **Demography**

The latest 2011 census sees a population of 53,012,456 living in England. With about a half of the UK's land, England is accommodating 83% of its total population (63,182,178) according to the 2011 census (Office for National Statistics, 2013a).

In China, there are 56 ethnic groups – 55 minorities (Zhuang, Uyghur, Hui, Manchu, Miao, Tibetan, etc.) and one ethnic majority (Han). In 2010, China had a population of 1,339,724,852, with 16.6% of them being aged 0-14, 70.1% 15-59 and 13.3% 60 and above (National Bureau of Statistics of China, 2011).

1.4.1.3 **Culture**

English culture is not much different from British culture as a whole. It has the characteristics of an island country, the features of a Christian-based society, and traces of a former world superpower. It has its own unique taste in literature, music, architecture, folklore, philosophy, etc. All of these have various connections with Christian-oriented beliefs and the priority given to freedom and democracy. England nurtured world's famous writers, such as Geoffrey Chaucer and William Shakespeare, and singer-songwriters, such as John Lennon and Paul McCartney (Keet, 2012).

Starting its evolution along two old rivers, the Yangtze River and the Yellow River, Chinese culture has its roots settled in Confucianism, Taoism and Buddhism. The three streams of philosophy particularly Confucianism have formulated the cultural phenomena that Chinese predominantly value harmony, family responsibility, education, good morality and diligence, yet historically generally lacking critical thinking and the courage to challenge authorities. After 1949, traditional culture in mainland China was altered by politics in many ways. The post-1979 era has seen a gradual decrease of such political influence over its traditional culture.

1.4.1.4 Politics

With London as its capital, England is a non-devolved constituent country within the constitutional monarchy and parliamentary system of the UK. England is directly governed by the British Parliament – one of world's oldest parliaments.

China is a centralised socialist country which is run by one party, the Communist Party of China. The electoral system is hierarchical, allowing limited democratic processes within the party and at the local village level.

1.4.1.5 Education

The Department for Education (DfE) in England is the national department responsible for education of people from basic to higher education. England provides compulsory education for people from the age of 4 up to 16. Schools are inspected by the Office for Standards in Education (Ofsted) and graded into four levels: outstanding, good, satisfactory and inadequate. English primary schools employ generalist teachers who teach almost all main subjects. Children often are grouped by ability within and/or across classes. There are often teacher assistants (TAs) working alongside class teachers to provide extra support to those with special education needs (SEN). During a school day, children often have to travel between classrooms for different subjects' learning according to their ability levels.

The Ministry of Education is in charge of education at all stages in China including higher education. China provides nine years' compulsory education to children (aged 6 to 15) from primary through lower secondary phases. Teaching Research Offices (教研室) at district (county) and municipal levels are responsible for school inspections, focusing particularly on the process and performance of teaching and learning. Nonetheless, performance details of primary schools are generally not open to the public, partially to prevent parents from selecting to live in certain catchment areas on purpose. Chinese primary schools employ specialist teachers and take the mixed-ability approach within and across classrooms in all subjects. Upon entrance pupils are randomly allocated to fixed classes where different subject teachers come and deliver lessons. There is no such role as a TA in Chinese schools – every teacher ought to teach independently.

Table 1.1 Overview of the educational phases in two countries

<table>
<tr><td colspan="2">China</td><td>Age</td><td colspan="2">England</td></tr>
<tr><td rowspan="3">Kindergarten</td><td>Lower Grade</td><td>3-4</td><td colspan="2">Preschool</td></tr>
<tr><td>Middle Grade</td><td>4-5</td><td>Reception</td><td rowspan="7">Primary</td></tr>
<tr><td>Upper Grade</td><td>5-6</td><td>Year 1</td></tr>
<tr><td rowspan="6">Primary</td><td>Grade 1</td><td>6-7</td><td>Year 2</td></tr>
<tr><td>Grade 2</td><td>7-8</td><td>Year 3</td></tr>
<tr><td>Grade 3</td><td>8-9</td><td>Year 4</td></tr>
<tr><td>Grade 4</td><td>9-10</td><td>Year 5</td></tr>
<tr><td>Grade 5</td><td>10-11</td><td>Year 6</td></tr>
<tr><td>Grade 6</td><td>11-12</td><td>Year 7</td><td rowspan="5">Secondary</td></tr>
<tr><td rowspan="3">Lower Secondary</td><td>Grade 7</td><td>12-13</td><td>Year 8</td></tr>
<tr><td>Grade 8</td><td>13-14</td><td>Year 9</td></tr>
<tr><td>Grade 9</td><td>14-15</td><td>Year 10</td></tr>
<tr><td rowspan="3">Upper Secondary/ Professional Institutions</td><td>Grade 10</td><td>15-16</td><td>Year 11</td></tr>
<tr><td>Grade 11</td><td>16-17</td><td>Year 12</td><td rowspan="2">Sixth Form / Further Education</td></tr>
<tr><td>Grade 12</td><td>17-18</td><td>Year 13</td></tr>
<tr><td colspan="2">Higher Education / Professional Institutions</td><td>18+</td><td colspan="2">Higher Education / Further Education</td></tr>
</table>

Table 1.1 shows a comparison between the two countries' schooling phases along the timeline of children's ages. Those cells shaded in light blue show the stages of compulsory education in each country. The bold age group, i.e. 9- to 10-year-olds at the primary phase, is the target age group of the EMT study. Pupils aged 9-10 years in China are attending primary Grade 4, whereas pupils at the same age are taught in Year 5 in England.

1.4.1.6 Economy and comparability

The first industrialised country on the planet, today the UK is still one of world's largest economies. Its average GDP per capita ($38,514) is ranked the 21st in the world (The World Bank, 2013).

China is now the world's second-largest economy after the US. Nonetheless, its GDP per capita ($6,076 in 2012) is ranked 86th out of 182 countries by the International Monetary Fund (Wikipedia, 2013).

Despite the disparity of GDP per capita between two countries, it is worthwhile to compare the relatively developed regions in China with the equivalent areas in England.

1.4.2 Southampton and Nanjing

1.4.2.1 General information

Sitting on the south coast of England, Southampton has a total area of 51.81 km^2 (Southampton City Council, 2013) and a population of 236,900 (Hampshire County Council, 2013). The proportion of children aged 0-14 is 16.4%, that of people aged 15-59 66%, and that of people being 60 or above 17.6%. The ratio between White British and other ethnic groups is 77.7% to 22.3% (*ibid.*). With a glorious maritime history, today Southampton is still home to luxury cruise ships and the fourth largest British port by tonnage. Other pillar industries include car and aircraft manufacture, petrochemicals, electronic engineering, etc. There are 79 state schools, 4 independent schools and 3 independent special schools in Southampton. Of all state schools, 21 are infant schools (for children aged 4-7), 16 junior schools (7-11), and 24 primary schools (4-11), 12 secondary schools (12-16) and 6 special schools (DfE, 2013). Like the case of other regions in England, primary school admissions give priorities to those who have siblings in the same school, live close to the school, are in special care, have faith in a specific religion, or went to a feeder school.

Nanjing is located at the lower reaches of the Yangtze River. Once an ancient capital of the country and now the capital of Jiangsu province, it has a total area of 6,598 km^2. In 2010, Nanjing had a population of 8.005 million, with 6.47 million living in the urban area. The ethnic group, Han, is 98.56% of the city's population the rest of which are from 50 minority groups. Nanjing has been a prosperous city since ancient times. There are 344 primary (for children aged 6-12), 160 lower secondary (12-15) and 95 upper secondary (15-18) schools in Nanjing (Nanjing Education Bureau, 2011). To ensure equal distributions of educational resources, it is a national policy that every child should be enrolled to the primary and lower secondary schools (i.e., the compulsory phase) of their catchment areas. Schools at the compulsory stage are not allowed to select prospects in any forms. In Nanjing, the Education Bureau does not publish or rank publicly the performance of any primary schools. The purpose is to prevent parents from rushing to buy properties in catchment areas of 'best' schools.

1.4.2.2 **Economy**

Southampton was ranked the 27th out of 45 cities in the UK in terms of competitiveness in 2013 (Huggins & Thompson, 2013). Its GVA (£52,800 per capita in 2013) was ranked 16th out of 63 cities (Centre for Cities, 2015). Overall, Southampton is representative of the medium-to-high group of cities in England socio-economically.

Nanjing's GDP per capita was ranked 11th out of 36 major cities (in population terms) in China (National Bureau of Statistics of China, 2013). Its pillar industries include electronics, petrochemical, steel and automobile (China Knowledge, 2014). Nanjing is representative of the medium-to-high level cities in China in socio-economic terms.

1.4.2.3 Comparability

The national and municipal data and information shows the international comparability of Southampton and Nanjing, in socio-economic terms, with each of them representing the medium-to-high level cities in its home country. There are, of course, differences in terms of culture, politics, geography and demography, which is unavoidable in international comparisons but will be taken into consideration in the interpretation of the research findings later.

1.4.3 The participating English schools

The English sample comes from three primary schools, EN-A, EN-B and EN-C. All Year-5 teachers along with their pupils from each school took part in the study.

Two of the schools were founded in the1950s and one in the 1990s. All three schools are junior schools which provide education to children aged 7 to 11 in Years 4 through 6 (i.e. Key Stage 2). All schools were graded as 'Good' according to their latest Ofsted reports up until 2012. In almost every classroom, there is a generalist teacher assisted by a TA (occasionally more than one). School EN-B is unique in that Teachers EN3 and EN4 do not have a TA whereas Teacher EN5 has two to three TAs in her low-ability class.

School EN-A has about 170 pupils organised in 2 classes of each year group. Both teachers (EN1 and EN2) in Year 5 have taken part in the study. The school has ability-based grouping within maths classes but does not have maths sets across classes. School EN-B has approximately 290 pupils and three classes per year group. Children are setted in maths both within and between classes. All three teachers (EN3, EN4 and EN5) from Year 5 take part in the study. School EN-C has about 520 pupils organised in five classes in each of the four years. Like School EN-B, the school employs ability grouping in maths within and between classrooms. All Year-5 teachers (EN6, EN7, EN8, EN9 and EN10) participate in the study.

A typical English school starts at 9 am and children have two periods (approx. 1 hour 40 minutes per period with a 15-minute break in between) in the morning and one period in the afternoon. A period usually consists of two lessons about two different subjects without a break. For example, the teacher might start an English lesson immediately after a maths lesson. Lunch usually starts at around 12 pm and lasts about 50 minutes. The formal school time ends at around 3:15 pm. Some schools have after school clubs every weekday, whereas others only on certain days, for example, Fridays. The class teacher will be responsible for all subjects' teaching except the foreign language, usually French. Some may also teach after school clubs. The classroom is normally also the teacher's office – typically a corner (about the size of 2-3 m^2) beside the IWB with a desk, a PC, some drawers, and other essential facilities.

1.4.4 The participating Chinese schools

The Chinese sample includes all Grade-4 maths teachers and their pupils from four primary schools in Nanjing. Schools CN-A and CN-B, both founded in 2007, are relatively newer than the other two Chinese schools. School CN-A has six classes in each of six grades (years) and approximately 1,200 pupils and about 80 members of teaching staff. School CN-B has six classes per grade (year) with about 1,400 pupils and 110 teachers. School CN-C was founded in 1934 and rebuilt on the same site in 2006. With a team of about 60 teachers, it caters for about 650 pupils from Grades one through six. Due to urban development and commercial expansion, the surrounding community is shrinking. Thus, the school has two classes in Grade 4. Founded in 1994, School CN-D has four classes in each grade and caters for slightly over 700 pupils with a team of 55 teachers. As a matter of policy, all schools take the mixed-ability approach.

Conventionally, a maths teacher – like teachers of other core subjects – teaches two parallel classes in the same grade (same content, same pace). In this study, the Chinese team includes three teachers (CN1, CN2 and CN3) from School CN-A, three (CN4, CN5 and CN6) from CN-B, one (CN7) from CN-C, and two (CN8 and CN9) from CN-D.

A typical primary school day in Nanjing starts at 8 am and ends at 3 pm. There are four sessions in the morning and two in the afternoon. Each session lasts 35 to 40 minutes. There is a 10-minute break between sessions 1 and 2, 3 and 4, 5 and 6 and a 25-minute break between sessions 2 and 3 for broadcast gymnastics and, in some schools, also eye exercises. Lunch starts at around 12 pm, and session 5 starts at about 1:30 pm. There are usually after school clubs offering non-academic activities on campus. Teachers teaching the same subject share an office where a desk, a PC and other essential facilities are provided for every teacher. Together, they form the teaching research group (教研组) of the specific subject. In many schools, teachers of the same subject are often further grouped by grade. For example, all Grade-4 maths teachers may share an office.

With the research rationale and research contexts in mind, we will now look at the research purposes that drive the study.

1.5 Research purposes

As will be reviewed in Chapter 2, the area of TER is bounded by national borders, has a strong quantitative characteristic, but lacks rich descriptions (Reynolds et al., 2011; Reynolds et al., 2002b). To make meaningful contribution to the area of TER and to children's education in mathematics worldwide, this study is intended to fulfil the research purposes as follows:

- To evaluate and correlate the quality of mathematics teaching with learning outcomes cross-nationally;

- To collect multiple voices on mathematics teaching and learning and promote international conversations on EMT;
- To connect *hard* measurement with *soft* views and throw light on the improvement of mathematics teaching and learning within and across countries.

These purposes then lead to the formulation of research questions and the design of research methods in Chapter 3.

1.6 Research questions

Motivated by the learning gaps between pupils across countries in international performance surveys, the study aims at measuring the teacher effects in two countries' mathematics lessons on pupil learning outcomes with internationally validated instruments and collecting multiple views regarding the effectiveness of mathematics teaching. Mixed methods were applied to collect and analyse data so as to answer the following questions:

RQ1 *What are the correlations between various teaching approaches and learning outcomes in mathematics across two countries?*

RQ1a What are the quantity and quality of mathematics teaching in classrooms across two countries?

RQ1b How do children perform in two standardised mathematics tests?

RQ1c How does teaching correlate with learning across mathematics classrooms in the two countries?

RQ2 How do different roles in and beyond classrooms perceive the effectiveness of mathematics teaching in the two countries?

RQ2a How does the researcher interpret the quality of mathematics teaching in the two countries?

RQ2b How do individual teachers view the effectiveness of mathematics teaching in general and that of the specific lesson they delivered for observation?

RQ2c How do groups of teachers view the effectiveness of two mathematics lessons – one from England and the other China?

RQ3 How does the rigid evaluation of teaching and learning (RQ1) connect with the diverse perceptions of different roles about the effectiveness of mathematics teaching (RQ2)?

The final section of Chapter 1 will provide a *map* of the thesis for readers to travel from chapter to chapter.

1.7 Structure of the thesis

To keep things simple and clear, the thesis is organised in a classical six-chapter format, with chapters sequentially being:

- Introduction (Ch.1)
- Literature Review (Ch.2)
- Methodology (Ch.3)
- Results & Findings I (Ch.4)
- Results & Findings II (Ch.5)
- Conclusions (Ch.6)

After this introductory chapter, **Chapter 2** offers insights into mathematics pedagogical theories and practices, international studies on maths learning and teaching, American literature on TER, British literature on TER, and other nations' literature on TER. The limitation of prior international comparisons and TER in various countries consolidates the research purposes and helps shape the research questions.

Chapter 3 follows the line of research purposes and questions and combines a set of mixed methods for collecting and analysing empirical data. The theoretical framework of the project is presented and helps guide the subsequent data collection and analysis processes. Then, detailed data-collection-and-analysis methods and procedures are illustrated, and ethical issues discussed, before the chapter conclusion.

Chapter 4 mainly focuses on the results and findings from the quantitative analyses of data. After the chapter introduction, sequentially comes the results and findings of the structured observations, teacher questionnaire, pupil questionnaire, standardised tests, correlations between teaching and learning, and pupil background differences.

Chapter 5 mainly concentrates on the results and findings from the qualitative analyses of data. It seeks to present rich voices of various roles in and beyond classrooms regarding the quality of mathematics teaching through unstructured observations, video-stimulated interviews with teachers and video-stimulated focus groups with teachers.

Chapter 6 addresses the research questions, relates results and findings of the study to relevant results and findings in the literature, points out the contributions, implications and limitations of the study and foresees directions for future enquiries.

CHAPTER 2 - LITERATURE REVIEW

Chapter overview

- *Chapter introduction*
- Mathematics pedagogy: theories and practices
- International studies on learning and teaching
- Early research on teaching effectiveness in the US
- The research on teaching effectiveness in the UK
- National studies on teaching effectiveness
- *Chapter conclusion*

2.1 Chapter introduction

This chapter helps to set the historical scene and current status of research on mathematics teaching effectiveness. It starts from theories and practices of mathematics pedagogy, passes by the historical development of international comparisons on learning and teaching across countries, goes through the early stages of research on teaching effectiveness in the US and in the UK, and finally arrives at national studies done in this area. At the end of this chapter, the conclusion is intended to reveal the gaps in the literature and explain the general relevance of this study to both research and practice.

2.2 Mathematics pedagogy: theories and practices

2.2.1 Introduction

Pedagogy is defined as 'the method and practice of teaching' in Oxford Dictionary (Oxford Dictionaries, n.d.). In other words, it is about *how* to teach. Studies on pedagogy therefore share the same vision – seeking the *best* way of teaching – with studies in teaching effectiveness research (TER) which will be reviewed in the last few sections of this chapter. Works on mathematics pedagogy so far are generally either theories or empirical studies driven by theories. In this section, we will mainly trace the roots of dominating pedagogical theories in school mathematics and reflect upon some of the most investigated pedagogical aspects.

2.2.2 Progressive teaching

Progressive education has its root in trendy thoughts and theories around the beginning of the 20th century – most specifically, John Dewey's thoughts on democracy and education (Dewey, 1916) and Jean Piaget's theory of cognitive development (Piaget, 1950; Piaget & Inhelder, 1969). Founder of pragmatism, Dewey emphasised the role of freedom in the classroom where children should be actively experiencing and acquiring knowledge through hands-on activities rather than passively accepting the knowledge transferred by the somewhat authoritative teacher. Piaget believed that children were born with existing ability to learn, that their learning – a combined result of children's physical growth and their experience – went through a fixed sequence of stages, and that therefore the classroom should be child-centred and individualised with minimal teaching input, so that every individual child's developmental needs would be met.

Deweyan and Piagetian theories had transformed the landscape of school education in the West in the 20th century and formed the heart and soul of progressive education which now retains its influence in Western schools. The fundamental change it brings to education is that the role of teachers in the classroom is to ensure that individual children are free to explore the intended knowledge independently with minimal formal teaching input. The underlying assumption of progressive teaching is that in this way children can progress through stages of learning with maximal freedom and minimal external intervention. Alternative names for progressive teaching include constructivist, discovery, problem-based, experiential and inquiry-based teaching (Kirschner et al., 2006).

In a progressive setting, teachers are not encouraged to deliver the knowledge directly, and children are expected to acquire the knowledge themselves through exploration at their own pace. Progressive teaching is normally time-consuming and results are not ideal. Research has shown either a loss of learning or a low level of knowledge acquisition amongst children taught in this way, although the low ability learners tend to enjoy the explorative learning process (Clark, 1982, 1989).

The UK has been at the forefront of the implementation of progressive education. In the late 1960s, the Plowden report team drew the blueprint of the country's educational reform by referring to Piaget's theory (Central Advisory Council for Education, 1967). Since then, progressive education, as appose to traditional education, has become the norm in British schools. The philosophy behind it even dominated the landscape of research literature on learning and teaching at the time. As Bennett argued, there was little on teaching but a growing body of literature on learning and "a few fairly crude theories of teaching, based not on the observation of practice, but loosely on American conceptions of democracy" (Bennett, 1987, p. 67).

In fact, at the primary school stage, explorative learning appears to work against the way the human brain processes information over the short and long terms

(Atkinson & Shiffrin, 1968). Learners can only store a small amount of information over a short period and their learning of new knowledge has to be built on prior knowledge that is stored in their long-term memory. In "unguided environments" (Kirschner et al., 2006, p. 80), the amount of information stored in long-term memory distinguishes sharply novice learners from expert learners in that the latter have rich experience and sufficient knowledge in their long-term memory which helps reduce their cognitive load and assists them to make sense of challenging and complex learning content, whereas the former encounter enormous difficulties in the absence of both internal and external guidance. Therefore, at the early stage of learning, external guidance – such as learning worked examples – is necessary, as Kirschner put it:

> It emphasizes the importance of providing novices in an area with extensive guidance because they do not have sufficient knowledge in long-term memory to prevent unproductive problem-solving search. That guidance can be relaxed only with increased expertise as knowledge in long-term memory can take over from external guidance.
>
> (Kirschner et al., 2006, p. 80)

The naïve intention of progressive teaching in school mathematics, physics or chemistry is to create an environment for novices to obtain knowledge in a way a mathematician, physician or chemist generates knowledge in their career lives. However, the premise for a professional mathematician or scientist to work in an explorative and *productive* way is that they have already accumulated a solid base of knowledge (internal guidance) in their fields, whilst this is not the case for a child (extremely talented excluded) in the classroom. "The practice of a profession is not the same as learning to practice the profession" (Kirschner et al., 2006, p. 83). In mathematics, teachers cannot largely rely on explorative and self-guided activities, since every bit of the body of mathematics knowledge was probably the result of one mathematician's or a number of mathematicians' lifelong exploration. Children have only one chance to learn a *curriculum* of school mathematics over a dozen of years or so.

Democracy and freedom should be advocated, but it cannot be assumed to be a magic bullet for the teaching of specific school subjects, such as mathematics. There has to be a scientific way of teaching mathematics which must be evidence based. Since children have only one chance, teachers and educators should not apply untested theories to them. Unfortunately, this happened and is still happening. Progressive education has been experimented in the West for quite a long time in the 20th century, parallel to which is the decline of pupils' understanding of fundamental knowledge in mathematics (e.g. see ICCAMS Maths, 2015).

2.2.3 Teaching and ability grouping

As a way of differentiating teaching, streaming or setting has also existed for a long time in schools in Western countries, such as the UK and the US. In England, streaming became the norm in the 1950s, and schools generally streamed pupils into different ability classes and taught them all subjects in a fixed class. Despite the reversing trend favouring the mixed-ability approach around the 1980s, ability grouping has retained its dominance in English schools. National policies have been every now and then reinforcing this climate. Notably, the year 1997 saw the publication of the White Paper, *Excellence in Schools*, which declared that setting was "worth considering in primary schools" and "should be the norm in secondary schools" (DfEE, 1997, p. 38). Since then, setting has been widespread up and down the country. In English schools, there are now different forms of setting that group pupils by ability either within classes, or across classes, or both within and across classes.

Widely practised, setting is however generally found negative towards children's self-esteem. Collecting data in six British schools, Boaler et al. (2000) found not only the negative influences of setting in maths on the emotion of pupils in both high-ability and low-ability sets (i.e. classes) but also the pedagogical challenge that maths teachers encountered as they strove to differentiate teaching. On the contrary, Kulik and Kulik's (1992) meta-analysis suggested positive effects of grouping by ability upon academic performance and denied the negative effects that setting might pose both academically and emotionally. Whilst the effect of setting on learning remains slightly debatable, ability itself does not seem to be the single major predictor in deciding which set a child might be assigned to, since the SEN and social background of a child also plays a considerable role in set assignment (Muijs & Dunne, 2010).

2.2.4 Teacher knowledge

Around the late 1970s, inspired by the Piagetian theory, researchers in maths education had sought to identify pupils' levels of mathematical understanding in specific areas, such as decimals, ratio, algebra (Hart et al., 1981). Formative assessment was considered helpful for teachers to understand pupils' mathematical thinking and to improve their own subject matter knowledge.

About teacher knowledge, Shulman's work is one of the few that educational researchers in maths and many other school subjects are very likely to recall first. In his seminal work, Shulman (1986) fiercely criticised the then standards for teacher evaluation and teaching and school effectiveness research in the US for their interest in teaching process and outcomes and their ignorance of the *content* which he categorised as (1) subject matter knowledge (SMK), (2) pedagogical content knowledge (PCK) and (3) curricular knowledge. Amongst the three domains of teacher knowledge, PCK was regarded as the most important, since the grasp of the SMK, for example mathematics, did not in itself mean the capacity of making others understand and grasp the knowledge equally well. Four more types of teacher knowledge were later added: (4) general pedagogical knowledge, (5) knowledge of learners, (6) knowledge of educational contexts and (7) knowledge of educational purposes (Shulman, 1987).

In both research and practice in maths education, teachers' subject related knowledge has been widely emphasised and explored. Through task-based interviews, Ma (1999) compared American and Chinese maths teachers' understanding of basic subject matter and pedagogical content knowledge and found that Chinese teachers not only showed a better understanding of problem solving procedures (usually also with several alternative solutions) but also generated richer and deeper accounts of the underlying concepts and better ways of teaching them than American teachers.

In the US, empirical evidence suggests that teachers' mathematical content knowledge poses significant impact on pupils' mathematics performance (Hill et al., 2005). In the UK, studies have shown the British trainee teachers' lack of SMK and PCK in teaching mathematical topics, such as subtraction, gradient and geometry (Jones et al., 2002; Rowland, 2010). Because of the lack of competence in mathematics and therefore the dearth of confidence in the subject themselves, many English primary teachers are not confident in *teaching* the subject. A case study suggests that teachers may become more confident after taking part in continuing professional development (CPD) programmes that aim at improving their subject matter knowledge (Hodgen & Askew, 2007). Another study (Howe et al., 2015) found higher learning gains in the classes of teachers who had participated in CPD programmes focusing on PCK in rational numbers. The study claimed that the effect appeared to be greater over a longer term.

Nonetheless, there was a lack of systematic analyses of the whole teaching process.

Overall, research efforts so far are still more related to SMK than PCK, and research findings available seem fuzzy and cannot thoroughly address the questions Shulman posed three decades ago:

> How does the successful college student transform his or her experience in the subject matter into a form that high school students can comprehend? … How does the teacher prepare to teach something never previously learned? How does learning *for* teaching occur?
>
> (Shulman, 1986, p.8)

Despite the insufficient output over the past few decades, research on maths PCK has broadened its vision to throw the spotlight on both teachers' understanding (internal) and their teaching processes and outcomes (external). This is applausive for at least two reasons: (1) the teaching process is the very means by which teachers' PCK is put into action; (2) the outcome of teaching and learning is the touchstone of teachers' capacity in communicating the subject content to pupils in a comprehensible manner.

2.2.5 Teaching with coherence

Teaching with coherence is another element that has been repeatedly discussed in work on maths pedagogy over the recent years (Ma, 1999; Schmidt et al., 1996; Shimizu, 2007; Stigler & Hiebert, 1999). A coherent lesson is expected to build close connections between segments of a lesson and between a series of lessons. Such connectedness should exist at the conceptual level rather than merely the surface level. Thus, strong subject matter and pedagogical content knowledge seem to be crucial for a teacher to teach coherently. Coherence is one of the things that maths teachers pay particular attention to in East Asian countries, such as China (Chen & Li, 2010) and Japan (Sekiguchi, 2006). Cross-national studies suggest that maths teachers from East Asia put more emphasis on the conceptual coherence of teaching (Cai et al., 2014) and deliver maths lessons in a more coherent manner than their American counterparts (Stigler & Hiebert, 1999; Stigler & Perry, 1988).

2.2.6 Teaching with variation

Like many other teaching theories, teaching with variation is built on various learning theories (Marton & Booth, 1997). At the heart of this teaching theory are two aspects of teaching: (1) teaching *concepts* in varying forms (representations) and (2) teaching *procedures* in varying ways (solutions). The idea is that, through teaching with variation, prior knowledge may be consolidated in new situations and new forms, and the essence of new knowledge may be grasped thoroughly in rich and diverse connections with prior knowledge (Gu, 1994). By presenting prior knowledge in varying forms or situations, teachers attempt to scaffold the learning process of new knowledge. Through comparing different solutions to one problem, learners are expected to be able to absorb the essence of specific knowledge and at the same time eliminate misconceptions about it. Teaching with variation is widely accepted and applied by maths teachers from East Asia, such as China (Huang, 2002; Huang et al., 2006), Korea (Park & Leung, 2006) and Japan (Stigler & Hiebert, 1999).

2.2.7 Teaching with reflection

Reflection is regarded as a crucial way for pedagogical development of teachers. Schön (1983, 1987) proposes three levels of reflection: knowing in action, reflection in action and reflecting on reflection in action. Both research and practice communities have been promoting reflection amongst maths teachers. In initial teacher education, pre-service teachers may improve their reflective skills and form reflective habits through reflecting on lessons delivered by their peers and themselves during placement, which over all may benefit their future teaching practice (Stockero, 2008). Whilst reflection in teaching maths may generate pedagogical change (Walshaw, 2010), well-established reflective communities are claimed to be helpful in cultivating reflective capacities amongst maths teachers (Potari et al., 2010).

2.2.8 Conclusion

A dominant theory behind pedagogy in general and maths pedagogy in specific in the West is progressive education which offers few options for formal teaching. A beautiful ‘dream’ behind this is that children could learn best if provided with plenty of opportunities to explore the content independently. Piaget's theory of developmental stages has apparently provided a rationale for ability grouping in schools in England and other countries.

Research in maths pedagogy focuses more on factors within teachers' mind (thinking and knowing) than factors without (teaching). It is thus largely unanswered how to teach the subject matter knowledge well and why. Almost all works on maths pedagogy are either in themselves theoretical or empirical but deeply rooted in certain theories. Empirical studies on maths pedagogy are mostly qualitative in nature, with quantitative data seldom collected. There were indeed several large-scale studies, but they focused explicitly on learners' mathematical thinking (e.g. Hart et al., 1981; Hodgen et al., 2014). Evidence gathered in these studies was more about learning than teaching. Studies that did focus on the teaching process were often limited to a specific teaching factor, such as subject matter knowledge, differentiation, variation and reflection. In particular, studies claiming to investigate teacher PCK often turned out to be an exploration of teachers' explanations or classroom dialogue of subject matter knowledge. Subject related *pedagogy* is still the "missing paradigm" – as Shulman (1986, p.6) put. In maths education research, the 'black box' that encloses the connections between teaching and learning is largely undisclosed.

Last but not least, there is also a lack of attempts in seeking international evidence for better pedagogy in mathematics, given the fact that mathematics itself has long gone international as a science, like the discipline of neuroscience – a discipline that has close connections with teaching and learning. Over the past half century, a considerable number of international studies had focused on maths learning outcomes, whereas only a small number of them had stepped into classrooms and systematically investigated the teaching process of the subject cross-nationally. The following section draws upon this part of literature.

2.3 International studies on maths learning and teaching

2.3.1 Introduction

International comparative studies consistently reveal a performance gap in mathematics between the East and the West (Mullis et al., 2008; Mullis et al., 2004; Mullis et al., 2000; OECD, 2001, 2010). Most recently, the Programme for International Student Assessment, i.e. the PISA 2012 (OECD, 2013) showed that the 15-year-olds from Shanghai, China had been ranked atop the league table in Mathematics, Science and Reading out of 65 countries and economies, while their peers from the UK achieved 26th, 21st and 23rd respectively in the three assessed subjects. In Mathematics, Shanghai pupils achieved a mean score of 613, whereas British pupils attained 494 on average, lagging behind their peers from Shanghai by 119 points.

The assessment disparity motivated various groups of people, including educational researchers, to find the reasons for the learning gap (e.g. Stevenson &

Stigler, 1992). Some examined the social-cultural factors (e.g. Chen & Stevenson, 1995), some investigated the degree of parental involvement (Cai, 2003; Cai et al., 1999; Cao et al., 2007), others compared the curricular content that pupils were expected to grasp (e.g. Ruddock et al., 2008), and some explored the nature of teaching and learning in the classroom (e.g. Schmidt et al., 1996; Stigler & Hiebert, 1999).

In mathematics education and, of course, in many other subjects, the central issues concern what should be taught and learnt, what is actually taught and learnt, and what is attained at the end. The team of the Third International Mathematics and Science Study (TIMSS) divided the educational curriculum into three fundamental dimensions: ***the intended curriculum***, ***the implemented curriculum***, and ***the attained curriculum*** (Schmidt et al., 1996, p. 17). Subsequently, ***the potentially implemented curriculum*** was added in between the original first and second dimensions, making it a four-dimensional conceptual framework(Schmidt et al., 2001). *The intended curriculum* consisted of the aims and targets that the pupils were expected to grasp, while *the potentially implemented curriculum* stood for the curricular materials and textbooks. *The implemented curriculum* referred to the process in which the intended and the potentially intended curricula were carried out at the level of schools and particularly the level of classrooms. *The attained curriculum* was what pupils had actually grasped and ultimately achieved in the end as part of their own knowledge. It can be tested through various assessments at the classroom level, school level, national level or international level. Research on teaching is a necessary approach in finding causes for differences of pupil attainment (the attained curriculum), as it is teachers who implement the curriculum in the classroom and provide children *opportunities to learn* the intended curriculum.

With regard to the four categories of curricula, international studies have been done in the comparisons of curricula, textbooks and curricular materials, teaching practices, and learning outcomes, respectively. Given the aim of the study as evaluating teaching practices in mathematics across two countries, China and England, the foci of this section are on international comparisons of *learning outcomes* and *teaching practices* particularly in mathematics.

2.3.2 International comparisons of learning outcomes

Since the first international assessment in the 1950s, there have mainly been three groups of cross-national assessments conducted by three different organizations: the International Association for the Evaluation of Educational Achievement (IEA), the Educational Testing Service (ETS) and the Organisation for Economic Co-operation and Development (OECD). England has participated in most of these assessments, while mainland China has taken part in just three of them: the second International Assessment of Educational Progress (IAEP 2) organised by ETS (Lapointe et al., 1992) and PISA 2009 and PISA 2012 by OECD (2010, 2013). Nonetheless, several other Chinese-oriented societies, such as Hong Kong, Taiwan and Singapore, have participated in most of the studies, generally ranked in the top five.

The IEA's International Studies

The International Association for the Evaluation of Educational Achievement (IEA) was the first organization to initiate mathematics achievement comparisons across nations. IEA (2012) was founded in 1958 and has subsequently conducted more than 30 research studies comparing pupil achievement not only in mathematics and science but also in other subjects across the world. The organisation's original motivation was to evaluate both the educational inputs and the educational outcomes. Its primary founders saw the educational world as a natural laboratory that could trigger optimal learning outcomes in children. In order to identify crucial elements that significantly influenced education, various factors were examined in relation to pupil achievement at the international level.

The First International Mathematics Study (FIMS) was conducted between 1963 and 1967, with 13-year-olds and pre-university pupils from twelve countries being comparatively assessed (Husén, 1967). The FIMS found that *the opportunity to learn* was the most significant predictor explaining pupil-achievement differences. In 1980-1982, the Second International Mathematics Study (SIMS) was conducted to evaluate pupil mathematics achievement across 20 countries (Robitaille & Garden, 1989). In 1994-1995, the Third International Mathematics and Science Study (TIMSS) tested more than 50 million pupils from 46 countries

(Mullis et al., 1997). Since 1995, the IEA has been conducting TIMSS in grades four and eight in a four year cycle in 1999 (known as TIMSS repeat) (Mullis et al., 2000), 2003 (from 2003 on known as Trends in International Mathematics and Science Study with the same acronym) (Mullis et al., 2004), 2007 (Mullis et al., 2008) and most recently in 2011 (Mullis et al., 2012).

The ETS' IAEP 1 and IAEP 2

As a successful private organization in the United States, the Educational Testing Service (ETS) has, since 1983, managed the National Assessment of Educational Progress (NAEP), which was and still is the largest national organisation in the US assessing pupil progress in a number of subjects over time. Thereafter, ETS has, alongside representative organizations from participating countries, coordinated two international assessments in mathematics and science; entitled International Assessment of Educational Progress (IAEP), they were aimed at testing pupils' knowledge and examining crucial factors influencing pupil performance and attitudes to learning these two subjects.

In 1988, the first study, IAEP 1 (Lapointe et al., 1989), assessed 24,000 13-year-olds in mathematics and science in twelve educational systems, and subsequently 175,000 pupils aged 9 or 13 from 20 countries were tested in 1990 in IAEP 2 (Lapointe et al., 1992). Of the two assessments, China only participated in the second with 1,650 13-year-olds assessed and ranked in the top place in mathematics in their age group. England took part in both of the assessments ranking 9th out of 12 countries in IAEP 1, and 11th out of 14 countries in the pupil group aged 9 and 11th out of 20 countries in the 13-year-old group in IAEP 2. The two countries' ranking details are shown comparatively in Table 2.1.

Table 2.1 Two countries' mathematics rankings in IAEP 1&2

IAEP		*China*	*England*
IAEP 1	12 educational systems	-	9th
IAEP 2	Age 9 (14 countries)	-	11th
	Age 13 (20 countries)	1st	11th

Source: Lapointe et al., 1989; Lapointe et al., 1992.

The OECD's Assessments – PISA

From the beginning of the new millennium, the Organisation for Economic Co-operation and Development (OECD) has been conducting a three-yearly survey, the Programme of International Student Assessment (PISA), in both its member countries and partner countries. The aim of PISA is to assess the extent to which 15-year-olds in different countries, at the end of their compulsory education, have acquired sufficient knowledge and have been able to apply such knowledge in their daily life. The assessed subject areas include reading, mathematics and science. England has participated in all four waves of PISA, whereas China has only participated in the latest, PISA 2009 (OECD, 2010) and PISA 2012 (OECD, 2013). Table 2.2 compares the participating and ranking status of the two countries' 15-year-olds.

Table 2.2 Two countries' mathematics rankings in PISA

PISA	*Shanghai, China*	*the UK*
2000 (32 countries)	-	8th
2006 (57 countries)	-	25th
2009 (65 countries)	1st	28th
2012 (65 countries)	1st	26th

Source: OECD

2.3.3 International comparisons of teaching practices

Besides the comparisons of pupil achievement across the world, there are also studies investigating teaching practices across countries. The two IEA's TIMSS Video Studies (Hiebert et al., 2003; Stigler et al., 1999), the OECD's Teaching and Learning International Survey (TALIS) (OECD, 2009a) and the Learner's Perspective Study (Clarke et al., 2006) are reviewed.

The IEA's TIMSS Video Studies

In parallel with the international assessment of the Third International Mathematics and Science Study (TIMSS), two video-based lesson studies have been conducted, firstly in three countries—Germany, Japan and the US in 1995 (Stigler et al., 1999), and secondly in seven countries—Australia, the Czech

Republic, Hong Kong, the Netherlands, Switzerland, Japan and the US in 1999 (Hiebert et al., 2003).

Among various findings, it is worth noting that, in the TIMSS 1999 Video Study (ibid.), Japanese pupils spent 74% of their lesson time in making decisions on how to use procedures instead of simply following the teacher's guidance and carrying it out. With regard to the types of problems that teachers asked pupils, Japanese teachers raised a larger proportion (54%) of problems in facilitating pupils to make mathematical connections or investigate mathematical relationships. Nonetheless, teachers in Hong Kong allocated more (84%) problems to pupils to get them familiar with the problem-solving procedures.

All countries applied a large amount of lesson time to mathematical work, typically through the approach of problem solving. However, despite surface similarities, there were considerable differences. It was through solving problems that the eighth graders in Japan were introduced to new knowledge. In addition, Japanese lessons also featured more whole-class teaching and longer time per task for pupil seatwork than other countries' lessons. In contrast, Dutch teachers allocated a greater percentage of time for their pupils to work on a set of problems during individual seatwork.

There was also a difference in the emphasis of lessons between countries. Despite the fact that both allocated 76% of lesson time to new knowledge and 24% of lesson time to reviewing prior knowledge, Japanese and Hong Kong lessons had significant differences. Japanese teachers gave more attention to the introduction of new knowledge, whereas Hong Kong teachers focused more on pupils' practice of new knowledge.

Lesson clarity and lesson flow were also taken into consideration. High lesson clarity and low lesson interruption were found in the classrooms of the Czech Republic, whereas the Dutch classrooms were found to have low clarity and frequent lesson interruption.

In addition to the main findings provided by the TIMSS Video Studies' formal reports, there have been a number of follow-up analyses conducted both by members of the TIMSS research team and by many other researchers.

For example, prompted by teachers' understanding of the latest reforms from the TIMSS Video Study and Japanese lesson study, Hiebert and Stigler (2000) proposed a research and practice system for the improvement of mathematics teaching in the USA. Nevertheless, what teachers thought could hardly explain the real causal effect relationship behind the legend of top performing countries. It is apparently dangerous to directly *borrow* such understanding from abroad and base an innovation at home on it without proving it works.

In another example, based on the TIMSS 1999 Video Study data, Leung (2005) reanalysed the characteristics of mathematics lessons from Japan and Hong Kong.

The secondary analysis found that pupils were exposed to more instructional content in East Asian countries, though they talked less than their peers in the West. In addition to the highest proportion of mathematical language used in mathematics problems, East Asian pupils were given a longer time to solve problems and more opportunities to do proofs, and they were more likely to be engaged in the lessons. Moreover, Hong Kong teachers covered more advanced content in a more coherent manner and with more developed presentations. Nevertheless, how such instructional differences had impacted on East Asian learners' achievement was not further investigated.

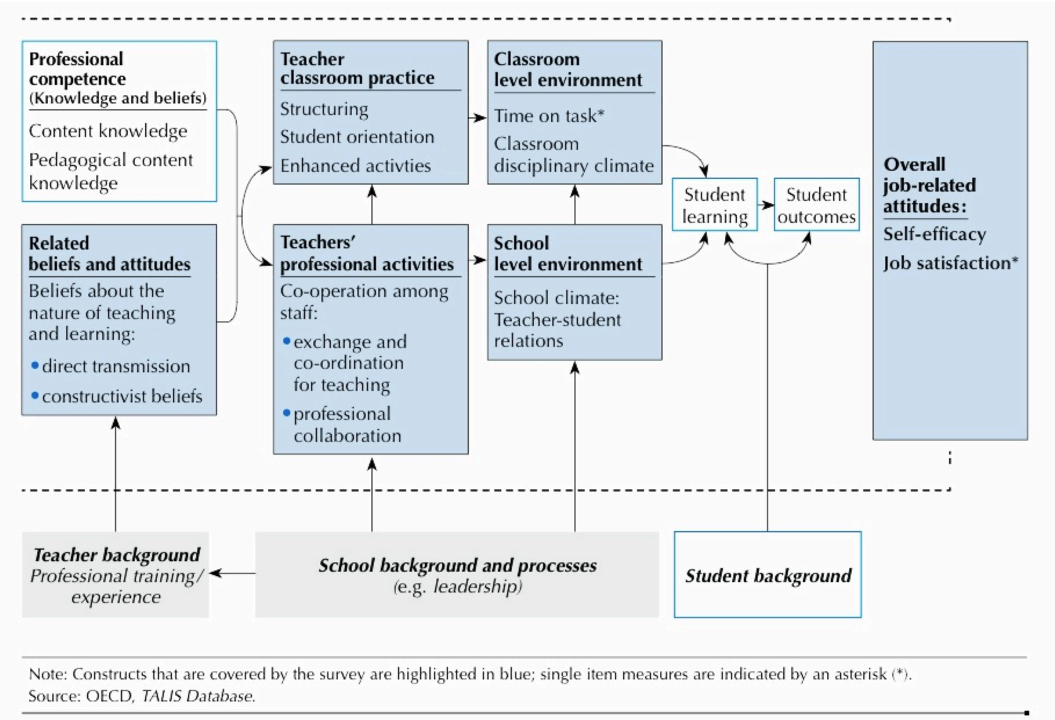
There are many other examples where TIMSS Video Studies' data have been either utilised to compare with lessons from countries other than those participating (e.g., Rossella & Alessandra, 2005) or reanalysed afterwards (e.g., Ilonca, 2001; Jacobs et al., 2006; James et al., 2005). However, it is systematic investigations into what worked across countries that was not brought to light in these studies.

The OECD's Teaching and Learning International Survey

To address the issues on the formulation of environments where effective teaching and learning can happen in different countries, the Organisation for Economic Cooperation and Development (OECD) has conducted its first round of Teaching and Learning International Survey (TALIS) (OECD, 2009a) and most recently TALIS 2013 (OECD, 2014). TALIS 2008 was based on data collected from 23 countries regarding "teachers' professional development; teacher beliefs, attitudes and practices; teacher appraisal and feedback; and school leadership" (OECD, 2009a, p. 3). There were 200 schools from each country and 20 teachers from each of the schools participating in the survey. The targeted population consisted of teachers and headteachers at the lower-secondary level, i.e., level 2 of the 1997 revision of the International Standard Classification of Education. Teachers and headteachers took part in the survey, filling in two different questionnaires, either on paper or on line.

In the TALIS 2008 Teacher Questionnaire (TQ), system-level and pupil-level variables did not draw much attention, whereas school-level and classroom-level variables were systematically surveyed (see Figure 2.1). In the area of teacher beliefs and attitudes, investigations were mainly done around the aspects of beliefs about the nature of teaching and learning, classroom teaching practice, co-operation among staff and job-related attitudes, self-efficacy, job satisfaction and so forth. About teaching practices, questions were asked with regard to time on task, classroom climate, classroom activities, lesson orientation, teachers' knowledge and so on.

In the domain of teacher beliefs, the study found a positive correlation between classroom disciplinary climate and teacher efficacy. In general, teachers across the countries believed that they made a difference. More than 90% of teachers from all countries believed that they made a great difference in their pupils' lives. It was also found that teachers in all countries expressed strong attitudes towards putting great emphasis on structured instructions and pupil-oriented activities. These were more applied by female teachers, whereas male teachers more tended to perceive teaching as the direct transmission of knowledge. In all countries, teachers generally allocated 80% of the lesson time to teaching and learning. However, teacher-pupil relationships varied dramatically across countries, except for Norway where the relationship was well above the international average level and where within-country differences were quite small. In practice, teachers in all countries were likely to set learning targets, make brief revision of previous lessons, give feedback to homework, examine pupils' understanding, etc.



Source: OECD, TALIS 2008 (2009a, p. 91)

Figure 2.1 TALIS framework for analysing teaching practices & beliefs

In terms of teaching, teacher questionnaires in TALIS 2008 and the subsequent TALIS 2013 focus intensively on teacher self-efficacy and teaching practices, offering insights into what teachers think they are doing across countries. Nonetheless, teachers' self-reported efficacy can only provide one-sided information. The results of TALIS survey could have provoked even deeper thoughts if there were extra data collected through other methods, such as classroom observations which would help reveal what teachers are actually doing in contrast with what they claim to be doing in the survey, focus groups which could lead to interactions between teachers on what counted as more effective lessons, and so on.

The Learner's Perspective Study

Through international collaboration, Clarke et al. (2006) conducted the Learner's Perspective Study (LPS) with researchers from twelve countries: Australia, China, the Czech Republic, Germany, Israel, Japan, Korea, the Philippines, Singapore, South Africa, Sweden and the US. Each country team chose three teachers as competent teachers on local standards. Given the way of selecting teachers, it was not LPS researchers' intention to sample representative teachers and typical lessons in each country for cross-national comparisons.

The research design was qualitative in nature. For every teacher, a sequence of ten consecutive lessons taught by each teacher participant was observed and video-recorded. For every classroom, three questionnaires were given to the teacher, and a post-lesson interview with the teacher and another with two focus pupils were conducted to collect their accounts on lesson events, using lesson videos as stimuli.

The reason LPS researchers did not intend to collect national samples and insisted on collecting in-depth qualitative data was a result of their critical reviews on the research methods and findings of the two TIMSS Video Studies. LPS researchers (Clarke et al., 2007) criticised the points that TIMSS Video Study researchers held regarding the focus of comparisons and the levels at which variables should be chosen and compared. They attempted to test the extent to which the national teaching patterns their TIMSS counterparts found would work on the LPS data from the US, Germany and Japan. Then, they employed the TIMSS Video Study lesson codes generated from lessons in the three countries and applied them to the LPS data set of the three countries. These codes were claimed to be not identifiable in almost each LPS lesson and no evidence suggested the existence of the national patterns that TIMSS researchers found. Conversely, there is a strong argument in the LPS report that lessons varied considerably within and between individual teachers' classrooms in the three countries (Clarke et al., 2007). They thus argued that, in international research on teaching, it would be more sensible to compare lesson events than to compare lesson patterns/scripts on which the two TIMSS Video Studies were focused.

Moreover, the TIMSS Video Study researchers (Givvin et al., 2005) reanalysed the TIMSS 1999 Video Study data and characterised each country's lessons in three dimensions: Purpose, Classroom Interaction and Content Activity. This aroused further debates from the LPS researchers over the undue simplicity of the TIMSS Video Study researchers' approaches. They argued that the TIMSS Video Studies' findings were based on "over-inclusive codes" (Clarke et al., 2007, p.283) which increased the international adaptability of the coding scheme but decreased its discrimination in reflecting inter-class variation. Thus, the LPS researchers decided to focus on "the location of the lessons in the classroom of competent teachers" and "learning outcomes provided in post-lesson interviews" (ibid., p.281). The former led them to observe a sequence of each teacher's lessons on a topic rather than one lesson and interpret connections between lesson events and teaching goals. The latter was collected in video-stimulated interviews with focus pupils.

Teaching studies conducted by the two teams of researchers on both sides have their strengths and weaknesses. TIMSS Video Study researchers tried to see things over holistically, whereas LPS researchers scrutinised lessons too closely therefore losing the opportunity to capture bigger pictures. Other TIMSS Video Studies' strengths and weaknesses have been discussed formerly. For the Learner's Perspective Study, its advantages lied in the application of innovative methods in research on mathematics teaching and learning and the rich accounts of the insiders' (i.e. teachers and pupils') thoughts. However, a heavy emphasis on the insiders' views prevented research findings from being thoroughly synthesised among countries. There was a severe lack of quantitative measurements to supplement with the multiple perspectives collected, which made the study less scientific than it could have been.

2.3.4 Conclusion

This section reviewed large-scale studies conducted in two domains of international comparisons: the comparison of learning outcomes and the comparison of teaching practices, which are correspondingly consistent with the two fundamental variables in the research on teaching effectiveness – pupil achievement and teacher behaviours – as will be reviewed later on.

The international assessments organised by the IEA, the ETS and the OECD over the years consistently revealed that pupils from a Confucian-heritage-cultural (CHC) background (including Taiwan, Hong Kong, and mainland China) steadily outperformed their peers in the West. Although pupils from mainland China only participated in three of those assessments, they performed the best and were ranked at the top on all occasions (Lapointe et al., 1992; OECD, 2010, 2013).

The second part of the review in this section had reflected on studies from three groups of large-scale international comparisons on mathematics teaching practices: the two IEA's TIMSS Video Studies (Hiebert et al., 2003; Stigler et al., 1999), the first TALIS report by the OECD (OECD, 2009a), and the Learner's Perspective Study with Hong Kong as an exemplar of China (Clarke et al., 2006; Mok, 2009).

The number of international studies on the teaching practices of mathematics is relatively small compared with that on pupil performance in the subject and very few of the large-scale comparisons reviewed involved teachers from mainland China. There is clearly a dearth of cross-cultural studies on teaching practices involving China, particularly given the results that Chinese pupils topped the international league tables every time they participated. Moreover, few of international comparative studies have attempted to measure and correlate learning outcomes and teaching practices in one study. Nevertheless, from the very beginning of teacher effectiveness research, the impact of teaching on learning has always been the main research interest of researchers in this field. The following section will serve to review early teacher effectiveness research literature originated in America in the 1970s and 1980s.

2.4 Early research on teaching effectiveness in the US

In the early 1970s, the development of human behaviourism theory contributed to and motivated research towards a concentration on whether teachers were teaching effectively, according to the developmental process of children's cognitions and the corresponding learning procedures in their minds. Effective teachers are the ones who, at the least, are able to successfully improve pupils' learning outcomes in the subject matter. In some sense, teaching and learning are regarded as a cause-and-effect pair. Since the core aim of teaching is to enhance pupil learning in specific subjects, it is essential to understand how learning is processed in children's minds before one starts to consider the most effective way of teaching.

Like many of their contemporaries, Rosenshine and Stevens (1986) linked the process of teaching to research on 'human information processing' (p. 378) so as to understand the ways in which teaching worked. According to information processing theory, human brains are limited to and are only able to effectively cope with a certain amount of information each time (Becker, 1978; Miller, 1956). Such a feature of human learning suggests that new knowledge needs to be broken into small segments and taught in small steps (Rosenshine & Stevens, 1986). Furthermore, new knowledge cannot be grasped and stored in a stable manner unless learners keep practising it to an overlearning degree so that it can be transferred from temporary memory to long-term memory, providing sufficient available space for more knowledge to come in (Gagné, 1985; LaBerge & Samuels, 1974). In different subjects, various layers of knowledge are constructed in different ways. Overlearning words enables pupils to pay more attention to the comprehension of reading materials, while overlearning basic skills, such as calculations in mathematics, enables pupils to fully concentrate on high-cognitive-level work, such as problem solving and mathematical reasoning.

Building on the principles of human information processing and the effective models developed by other researchers (Gagné, 1970; Hunter & Russell, 1981), Rosenshine and Stevens (1986, p. 379) suggested a general model of effective instruction which contained six components: (1) daily review and checking homework, (2) presentation, (3) guided practice, (4) correctives and feedback, (5) independent practice and (6) weekly and monthly reviews. This structure echoed with different effective teaching models identified by other researchers (Good & Grouws, 1979) and showed its consistency with effective teacher behaviours found in other studies (e.g., Fortune, 1967; Soar, 1977). These models commonly emphasised the necessity of revising prior knowledge before inputting new knowledge and the need to develop teaching steps according to the internal structure of various cognitive levels of knowledge.

Around the 1970s and 1980s, there were hundreds of studies done on teaching effectiveness in the US. Researchers usually did pre- and post-tests of pupil academic performance and then linked the differences between pupil test results over time to teacher behaviours which were mainly measured through classroom observations or questionnaires. Many studies consistently found positive effects of certain teacher behaviours on pupil achievement. This section of literature review is therefore aimed at seeking the most identified effective teacher behaviours and building connections between those behaviours.

Firstly, the key definitions of teacher effectiveness featured in the literature will be retrieved; secondly, the developmental process of research methods and theoretical frames in teacher effectiveness studies in the 1970s of America will be briefly reviewed; and then the main discussion will address effective teacher behaviours identified: (1) *profound subject matter knowledge*, (2) *academic-oriented instruction*, (3) *high opportunity to learn*, (4) *more time allocated to direct instruction*, (5) *clear and structured presentation*, (6) *maintaining a brisk lesson pace*, (7) *prompting a moderate to high success rate*, (8) *active whole-class interaction*, (9) *proper use of teacher questioning*, (10) *appropriate praise and criticism*, (11) *assigning homework regularly with timely and clear feedback* and (12) *efficient classroom management*. Finally the identification of gaps in the literature will be presented and discussed as a conclusion to this section of the literature review.

2.4.1 The definition of teacher effectiveness

Teachers' teaching at the classroom level is seen as a process towards gaining certain expected products, i.e. learning outcomes. The more learning outcomes produced, the more effective the instruction (process) is. Research on the connections between the process and the product was called *process-product research* in 1970s America. Based on such ideas of effective teaching, earlier American teacher effectiveness researchers carried out a huge amount of classroom studies, mainly through lesson observations, correlating teacher behaviours with learning outcomes (Brophy & Good, 1986).

2.4.2 The methodological and conceptual development

Prior to the 1970s, there was a lack of "systematic research linking teacher behaviour to pupil achievement", and the existence of such a research gap was partly attributed to the fact that the cost of observations at that time was considerably high and also partly caused by previous concepts on teaching (ibid., p. 329).

In the era before the 1970s, researchers had already started seeking factors in teachers that would increase the quality of teaching and learning, but they were more interested in connecting pupils' achievement with teachers' personal traits than with specific teacher behaviours. Thus, Brophy and Good argued that the "early concern with teachers' personal traits led to presage-product rather than process- product studies" (1986, p. 329).

Thereafter, research forms were limited at quasi-experimental comparisons of the achievement of pupils taught in one way of teaching with the achievement of those taught through another teaching approach. There were just a few studies conducted in this domain (ibid.) among which apparently little difference was identified regarding pupil achievement between classes taught with different methods.

The systematic observation of teacher behaviours was triggered by the educational trend in facilitating good classroom climates and effective teaching in the 1950s and 1960s, and some modest findings were achieved (e.g., Flanders & Simon, 1969; Gage, 1965). However, at the same time, there was a concept calling for a shift of the research focus from teachers to the curricula, i.e. *what* to teach, on which some researchers put more emphasis than on *how* to teach (e.g., Walker & Schaffarzick, 1974). Moreover, early school effectiveness research also emphasised less the contribution of teachers than that of schools to the improvement of pupils' achievement.

Nevertheless, during the 1960s, significant progress in research conceptions and methodologies had also been made and gradually proved the process-product approach to be a valid way of researching educational effectiveness. New observation systems had been developed with new process variables identified. The number of observation systems had dramatically increased up to more than a hundred by the year 1970. Meanwhile, research had identified steady relations between particular teacher behaviours and pupils' learning outcomes (Rosenshine, 1971). Such correlations were predicted to possibly be working in nonlinear ways and dependent on differences of pupil ability and background. In addition, Rosenshine (ibid.) pointed out that several key factors, such as contexts, the subject matter, grade levels and terminologies applied in the measurement, might have a significant influence on research findings in this area. From the beginning of the 1970s up to the early 1980s, a great number of studies had been conducted and new variables examined in the area of process-product research on teaching and its impact on learning.

Overall, there were some events that contributed to the continuous development of research on teaching during the 1970s. Rosenshine and Furst (1973) called for the utilisation of direct observations in the area of teaching studies as well as systematic research on the descriptive-correlational-experimental loop to describe lessons with quantitative instruments, correlate observed variables with pupil achievement, and examine corresponding causes of specific effects in the correlational connection. Moreover, Dunkin and Biddle (1974) defined research on teaching and made distinctions between teaching studies and research in other educational areas. They also critically stressed the importance of evidence-based policy making, argued the inappropriateness of making educational prescriptions based on untested theories and pointed out the necessity of prescribing education based on reliable empirical findings. In addition, they emphasised the importance of contextual variables in research on teaching. Last but not least, the American federal agencies, such as the Office of Education and the National Institute of Education, recognised the importance of research on teaching and provided sufficient funding for numerous large-scale studies on the effectiveness of teaching, which provided a positive national climate for teacher effectiveness research (TER) to flourish.

2.4.3 Effective teacher behaviours identified

The great number of TER studies makes it feasible that the most frequently observed and identified effective teacher behaviours could be located and organised for further analysis and synthesis. In this subsection, those essential teacher behaviours will be presented, discussed and synthesised.

Profound subject matter knowledge

As research indicated, effective teachers generally demonstrated stronger subject knowledge and clearer understanding of how to put it into well-structured instructional practices than did less effective teachers (Tikunoff et al., 1975). This was consistent with the recognition of the importance of teachers' pedagogical content knowledge in Schulman's seminal work (Shulman, 1986, 1987) and also with Ma's (1999) findings that the gap in mathematics learning between pupils from two countries was partly attributable to Chinese primary teachers' more profound understanding of the subject matter than their counterparts in the US.

Academic-oriented instruction

More effective teachers were found to spend more time on academic activities and successfully facilitate pupils to engage in such activities longer than less effective teachers (Fisher et al., 1980). They were also good at briskly switching between lesson segments, therefore spending less time on transitions and discipline maintenance than their less effective peers (ibid.). Teachers who produced the largest learning gains devoted the maximum proportion of time to active academic teaching and seatwork supervision and the minimum proportion of time to managing pupil behaviours (Berliner & Tikunoff, 1976, 1977). Effective teachers were also good at tailoring appropriate subject content and academic tasks for pupils according to their current knowledge levels and readiness for new knowledge (Fisher et al., 1980).

In the Follow Through Evaluation Study, Stallings and Kaskowitz (1974) studied lessons in 108 Grade-1 and 58 Grade-3 classes taught by experienced teachers who were implementing one of the seven Follow-Through models. Data were collected regarding the amount of time allocated to various activities and the frequencies of teacher-pupil interactions. The results indicated that the teaching components derived in the Follow-Through models had a positive impact on pupils' standardised test scores. Overall, the pupils that gained more were the ones whose teachers used more time in active instruction of subject content and provided sufficient supervision to pupils during individual work when necessary. On the contrary, pupils that gained less were the ones whose time was largely spent on non-academic activities and whose work was completed independently with little supervision. Given the pupil attainment level before the study as below average, the findings also suggested that low-ability pupils might need more frequent instruction from their teachers so as to improve academic performance.

The California Early Childhood Education programme conducted by Stallings, Cory, Fairweather and Needels (1977) also obtained similar findings which coincided with those from the Follow Through Evaluation Study in that the quantity of teachers' academic instruction played an essential role in improving pupils' achievement.

Another study by Stalling and her colleagues (Stallings et al., 1978), on reading instruction in 43 secondary schools, again identified the extent of academic-oriented instruction as a key factor correlating with pupil academic performance. Negative factors included little teacher-pupil interaction, teacher classroom management (instead of academic instruction), pupil independent work without sufficient supervision and so forth.

High opportunity to learn

In classrooms where effective teaching could happen, besides sufficient orientation of teaching in academic content, another crucial factor is providing the highest opportunity for pupils to learn (Stallings et al., 1977). The teacher, as the initiator of such an opportunity, should not only ensure a large enough *quantity* of learning, such as a high coverage of curricular content, pupils' more time on task, etc., but also guarantee a high *quality* of such academic engagement. Pupils should be thoroughly engaged into the thinking and learning of the subject knowledge. Thus, an ideal opportunity to learn consists of two mutually dependent dimensions: *time allocation* and *academic engagement*, which cannot create change independently and can only work collaboratively on the improvement of learning (Arehart, 1979; Borg, 1980; Crawford et al., 1978; Fisher et al., 1980). In general, teachers delivering high-quality lessons tend to leave plenty of time for teacher-directed instructions, allocate less time to individual seatwork, and maintain equal supervisions to every pupil during seatwork, particularly at the primary phase.

More time allocated to direct instruction

Initially motivated by the interest in observing the amount of teacher talk, Flanders and his colleagues designed an observational instrument, the Flanders Interaction Analysis Categories (FIAC) (Flanders, 1970), in order to describe directness and indirectness of teaching in the classroom. Coding teacher behaviour with 10 behavioural categories in the FIAC system, the researchers found teacher talk occupied two thirds of classroom conversations and had positive effects on pupils' achievement. Moreover, indirect teaching was found to have more positive correlation with pupils' attitudes than with their achievement. Direct teaching was found to be more appropriate for teaching facts or skills, whist indirect teaching was found to be more suitable for teaching abstract reasoning or creativity. Since basic skills and facts were mainly addressed at the primary phase, direct teaching was much more appropriate for this stage of learning than was indirect teaching. Conversely, indirect instruction was much more suitable for triggering creative thinking at the secondary phase. Teacher-directed teaching might not, however, always work effectively in improving pupils' achievement without clear presentations and other effective elements to complement it.

Clear and structured presentations

In implementing a session, effective teachers were found to be more able to plan ahead and then present their teaching in a clearer and more structured way than others. Such clarity is mainly evident in two domains: teacher expressions and lesson structures.

Clarity in teacher expressions

A strand of research interest grew around the clarity of teacher presentations during lessons, with the main negative factors being vagueness of terms, digressions, discontinuity and saying “uh”. Smith and Land (1981) found pupils’ achievement scores were influenced and reduced when vague terms were added into well-planned lessons. In addition, in three out of four studies, they also discovered similar decreases of pupil academic performance when clear presentations had been mixed with digressions, for example, when teachers repeated certain words or added ‘I mean’ in between words of a sentence which was originally planned to be fluent. Evidence from Kounin’s study (1970) proved that lesson momentum could be interrupted by any discontinuity delivered by teachers, for example, in mentioning irrelevant content or introducing relevant content at an unsuitable time. Similarly, Smith and Cotten’s (1980) study suggested that a high level of content changes made to the original clearer presentation led to the reduction of pupil achievement. Teachers with a high level of teaching clarity were able to generate higher pupil achievement (e.g., Land, 1979; McCaleb & White, 1980).

Clarity in lesson structures

Apart from clarity in teacher expressions, some studies also focused on the correlations between the clarity of lesson structures and pupil learning outcomes. Applying both low-inference and high-inference ratings, Fortune (1967) observed and analysed trainee teachers’ presentational skills in teaching English, mathematics, or social studies in upper-primary classrooms. The findings clearly distinguished more effective teachers from less effective ones in that the former were more likely to start lessons with an overview or analogy, and reviewed or repeated old knowledge when necessary.

In the Missouri Mathematics Effectiveness Project (Good & Grouws, 1979) – a treatment study, an instructional model was introduced to teacher participants. This model consisted of five main domains: daily review, development, seatwork, homework assignment, and frequent reviewing, with more emphasis on active whole-class teaching. The implementation of the instructional structure with clarity proved to significantly improve pupils' achievement over several months in the classes taught with the model. Clark et al. (1979) also found a positive correlation between clearly structured lessons and an improvement in pupils' achievement. Furthermore, clear transitional signals between the components of lessons were also regarded as crucial in making lessons well structured and more effective.

Overall, clearer lesson structures coincided with the information-process theory in building connections between old and new knowledge and also between short-term and long-term memories by mapping revision, explanation, demonstration and assigning classwork and homework properly throughout the lesson. This means that the clarity of lesson structure has its effectiveness rooted in both empirical and theoretical areas.

A brisk lesson pace with a moderate to high success rate

The speed at which teachers teach varies considerably across lessons and also makes the levels of teachers' effectiveness significantly different from one another.

A longitudinal large-sample study, the Texas Teacher Effectiveness Study (Brophy & Evertson, 1976) followed 165 teachers' teaching through four consecutive years, paying attention to the correlations between teaching pace and pupil learning outcomes. It was found that pupils learned best when their teachers broke learning tasks into small and easy steps and then taught them at a brisk pace. Moreover, some differences were also identified between low-SES and high-SES pupils in terms of the corresponding teaching pace and how small the steps should be. Low-SES pupils should be taught in smaller steps with certain repetitions when necessary, while high-SES pupils could benefit more from a brisker instructional pace and in slightly larger steps.

In the Stability Analyses study, Good and Grouws (1975) collected data from lessons given by 103 third and fourth grade teachers in two consecutive years and observed the stability of teaching effects on pupil attitudes and achievement in all subjects. The results were statistically significant but at low levels. They then decided to focus on one subject, mathematics, by comparing the teaching effects of nine effective teachers with those of nine less effective teachers on their pupils' achievement. In a later study, the Fourth Grade Naturalistic study, with an extra group of 23 fourth grade teachers added to the original 18 teachers, Good et al. (1978) discovered that more effective teachers taught the curricular content at a brisker pace than their peers.

Overall, a brisk pace saves teachers' teaching time to address more new knowledge, and consequently exposes their pupils to more academic content in a session, which significantly increases their opportunity to learn (Good et al., 1983; Griffin & Barnes, 1986; Lampert, 1988).

Prompting a moderate to high success rate

In addition to a brisk pace, collecting data from 25 grade-2 and 21 grade-5 classes, researchers from the Beginning Teacher Evaluation study (Berliner et al., 1978; Fisher et al., 1978) also found that high success rates were positively correlated with pupils' learning outcomes. For the whole sample, on average, pupils achieved 50% high success, about 50% medium success, and 0-5% low success. Such effects varied slightly across different pupil age groups. The learning of younger pupils was promoted more by high success rates than that of older ones for whom a moderate level success rate was appropriate. For example, grade-5 pupils averaged 35% high success rates. Overall, it was found that pupil achievement was positively correlated with approximately 65-75% high success, 25-35% medium success and 0% of low success.

Active whole class interaction

Besides the effective behaviours discussed above, in a holistic view, effective teachers were found to be teaching more actively at the whole-class level and thus striving towards a high level of pupil engagement.

Good and Grouws (1979) conducted process-product research along with their colleagues in the 1970s and developed a mathematics teaching model called active whole-class instruction in mathematics. They found the teachers who implemented whole-class instruction could be either effective teachers or less effective teachers. Whole-class instruction, they noted, would only add value to the effectiveness of teaching if teachers implemented it in a more active way, with efficient classroom management and appropriate teaching skills. Likewise, in the Junior High study, Emmer et al. (1979) found that more effective teachers allocated more time to whole-class teaching and demonstration but less time to seatwork, taught more actively and aroused more class-level discussion.

In general, effective teachers conduct more interaction with the whole class rather than with individuals with whom they only interact in a brief but efficient way. Supervision to each pupil during the seatwork stage is evenly distributed. Effective teachers attempt to track their pupils closely in the whole class scope and ensure their thorough comprehension of the knowledge before sending them to seatwork in which re-teaching is not necessary and therefore less time is needed for completing the practical work than in less effective teachers' classes (Evertson et al., 1980).

Proper use of teacher questioning

Abundant evidence also consistently confirms the importance of teachers asking appropriate questions for improving learning outcomes.

More effective teachers are less likely than other teachers to ask questions to which pupils are not able to provide correct answers (Good & Grouws, 1977). On the other hand, there is also evidence for the curvilinear relationships between pupils' academic performance and percentages of correct answers they make regarding teachers' questions. Brophy's (1973) large-scale study found different optimal levels of correct-answer percentages for high-SES and low-SES pupils: the former progressed best when around 70% of their answers were correct; the latter did best when about 80% of their answers were correct. The types of questions teachers asked therefore mattered. Effective teachers tended to ask more direct factual questions so that their pupils were more able to respond quickly with correct answers. In the Follow Through Evaluation study (Stallings & Kaskowitz, 1974), teachers who utilised more close-ended questions were found posing greater positive effects on pupil academic performance than those who applied more open-ended questions.

More effective teachers ask their pupils more questions than teachers with lower effects (Emmer et al., 1979; Evertson et al., 1980). There are also positive correlations between the quantity of process questions and the level of pupil achievement. Teachers asking more process questions than product questions provide an essential opportunity for pupils to clarify their thinking in the class, which at the same time makes their thinking more observable to the whole class and the teacher. The teacher is then able to diagnose more accurately the level of pupils' understanding of certain points of knowledge and timely prescribe pupils with optimal 'treatment'.

Appropriate praise and criticism

More effective teachers are good at maintaining a balance between praising and criticising pupils when confirming or denying their academic or conduct performance. The appropriateness of praise or criticism is among the factors most investigated in the studies that belong to this historical part of literature.

In reacting to pupil responses or giving seatwork or homework feedback, teacher praise at a moderate level generally generates a positive effect on pupil attitudes but mainly produces a positive achievement change in low-SES classes; it is somehow irrelevant to or even shows a negative connection with pupil achievement in high-SES classes (Brophy, 1973). Nevertheless, teacher praise is unrelated to learning improvement when such praise is about pupils' good conduct. In addition, there are negative correlations between teacher criticisms or punishment about pupils' misconduct and learning gains. When coping with misconduct, effective teachers tend to simply react with commands or warnings instead of criticism or punishment which less effective teachers usually employ.

In responding to pupils' wrong answers, effective teachers tend to make brisk and clear judgement without hurting personal dignity and then gradually give clues to pupils so as to enable them to answer the question more independently (Anderson et al., 1979). In addition, in summarising the incorrectly-answered questions, effective teachers tend to explain the process rather than simply providing the correct answers, hence deepening pupils' understanding and lowering the possibility of their future repetition of similar mistakes. With the theory of the inverted-U curvilinear process-outcome relationships, Soar and Soar conducted five process-outcome studies in 55 classrooms of grades 3-6 (Soar, 1966), 20 grade 1 classrooms (Soar & Soar, 1972), 59 grade 5 classrooms (Soar & Soar, 1973; Soar & Soar, 1978), 22 grade 1 classrooms (Soar & Soar, 1973; Soar & Soar, 1978) and 289 primary grades classrooms (Soar, 1973). Six observational systems were applied in the five studies in which negative climates, such as teacher criticism and negative emotions, were found negatively correlating with pupils' performance, and positive climates, such as teacher praise and positive emotions, were found having no significant positive relation with pupils' achievement. Moreover, for low-SES pupils, positive effects become more positive and negative effects more negative.

Assigning homework regularly with timely and clear feedback

Effective teachers tend to assign, evaluate and provide feedback to homework more frequently than their less effective counterparts (Good & Grouws, 1979). Starting a lesson with a review of the previous lesson, effective teachers are likely to process such a review by giving feedback about the previous homework to the whole class and thus reconstruct the prior knowledge in a class-level discussion before introducing new content, which is in line with the information-process theory (Becker, 1977; Emmer et al., 1982; Good & Grouws, 1979; Reid, 1978-1982). However, the positive effect of homework cannot happen without the functioning of the complete set of teaching factors. Judging by the fact that pupils are already capable of dealing with the seat-work content similar to the intended homework, an effective teacher is able to anticipate pupils' readiness for the homework before assigning it (Good & Grouws, 1979).

Efficient classroom management

Effective teachers are good at maintaining the teaching and learning process in order, allocating time reasonably to various activities, interacting properly with pupils and building a positive relationship with them.

Efficient time management

More effective teachers are able to manage the class more efficiently because they are better at identifying and dealing with problems immediately or preventing them from happening. They are more focused on the academic targets and therefore allocate more (ideally all) time to academic activities than less effective teachers. For instance, in the Fourth Grade Naturalistic study, Good and Grouws (1977) found that effective teachers could manage their classes better than their less effective colleagues even when the class size was bigger. Furthermore, effective teachers spend less time either on the transition between lesson segments or in maintaining classroom order. Effective teachers also spend less time dealing with previous assignments, because their pupils have already grasped the knowledge from the previous lesson(s) thoroughly and firmly.

Moreover, effective teachers paced forwards through the textbook more briskly, covering 1.13 pages per day on average, while their peers with lower efficacy could only cover 0.71 pages every day in the study by Good et al. (1978). Consequently, it was summarised that the brisk pace in effective teachers' lessons was formulated by firstly starting with clearer presentations, secondly knitting together explanations and questions without densely utilising teacher monologue, thirdly asking factual questions and promoting immediate correct responses and fourthly explaining the process rather than simply giving correct answers when reacting to incorrect responses (ibid.).

Even supervision to each individual seatwork

Good and Grouws (1979) found that, in addition to less time on homework revision, there was less time spent on seatwork in effective teachers' lessons. Effective teachers were found to be good at allocating their attention evenly to each individual and sufficiently supervising every pupil at seatwork, while less effective teachers tended to leave their pupils to independent work without requisite guidance (ibid.).

The way of formulating teacher-pupil relationship

Effective teachers have more sense of professional responsibility and are more willing to interact with pupils (Brophy, 1973). They are more likely to build a business-like instead of personal relationship with pupils, are keener on their work and interact with their pupils more politely and happily than ineffective teachers.

2.4.4 Conclusion

TER studies, initiated during the 1970s in the US, form a profound knowledge base for succeeding researchers home and abroad to advance the field upon. The review of this part of literature sets the historical and theoretical background for the study and suggests certain gaps. With regard to the research approaches, the gaps are mainly: (1) quantitative methods dominated the research convention and qualitative methods were rarely utilised; (2) there is a lack of teacher voices in what works; (3) there is a lack of international dialogue regarding the effectiveness of teaching. Moreover, as Rosenshine and Furst (1973) called for in the early 1970s, there is still a lack of understanding on the extend to which effective teaching is bounded by certain factors, such as grade levels and subjects. Models of effective teaching for specific schooling phases and in various subjects, for example mathematics, are still fuzzy, awaiting more recent systematic studies to push the research boundaries and inform practice with updated evidence.

Following the review of TER in the US is a review on the historical development of TER in the UK.

2.5 Research on teaching effectiveness in the UK

2.5.1 Introduction

By the end of the 1960s, British education had been much influenced by the then popular idea of child-centred classrooms which were recommended by the 1967 primary review known informally as the Plowden Report (Central Advisory Council for Education, 1967). The idea was also strongly influenced by Piaget's (1962) child cognitive developmental theory published during the same decade.

During the same period, the concept of process-product research on teaching effectiveness was imported from the US to the UK. British researchers carried out a number of studies on teaching effectiveness (e.g., see in Reynolds, 1985) among which were historically six major large-scale studies investigating the effectiveness of teaching, applying the process-product research concept and collecting and analysing large amounts of data over time. In chronological order, the first large-scale teaching effectiveness study was the study of Observational Research and Classroom Learning Evaluation (ORACLE) conducted near the end of the 1970s by Galton and his colleagues (1980) in 58 classrooms and19 primary schools; the second was Mortimore et al.'s (1988) School Matters project; the third, published in the same year, was Tizard et al.'s (1988) Young Children at

Schools in the Inner City which followed 343 children through three school years; the fourth was Pollard et al.'s (1994) PACE (Primary, Assessment, Curriculum and Experience) study investigating the changes that the 1988 Education Reform Act brought to schools, teachers, young learners and their perceptions of the new set of academic assessment, SATs; the fifth was the Effective Teachers of Numeracy (ETN) project carried out by Askew and colleagues (1997) from King's College London with a sample of 90 teachers and over 2,000 pupils; the sixth and most recent one was the Mathematics Enhancement Project Primary (MEPP) conducted by Muijs and Reynolds (2003) in 36 primary schools in England and Wales. Although overall TER studies carried out in the UK were much fewer than those done in the US, most strands of evidence further supported what had been proved to be effective teacher behaviours in the American literature.

In the remaining part of this section, the research designs and methodologies that the five large-scale studies utilised will be reviewed so as to shed light on the research design and methodology of this study; and then the findings of effective teacher behaviours that emerged in and considerably overlapped with each other among the five studies and other TER studies in the UK will then be analysed, integrated and linked with one another where possible. At the end of this section, research gaps will be summarised and presented in terms of methodologies and findings, and indications for this study will be concluded.

2.5.2 The research designs and methodologies utilised

The ORACLE Study

As the first large-scale longitudinal observational study on teaching effectiveness in primary classrooms in the UK, the ORACLE study (Galton & Simon, 1980; Galton et al., 1980) was conducted in 58 primary classrooms, across three local authorities, over three years. Pupil learning outcomes were tested against standardized testing papers for each school year. A huge amount of information was collected through 47,000 observations of 58 teachers and 84,000 observations of 489 pupils.

The discussion on the factors in relation to teaching effectiveness was mainly spread around teacher-pupil interactions, such as teacher questions, teacher statements, audience and content of interactions, pupils' time on task, teachers' attention to individual pupils and grouping strategies. The study found that the interactive whole-class teaching correlated positively with pupil learning outcomes/gains; that on average pupils in England spent 2/3 of classroom time working on their own; that of all interactions of teachers with pupils, 71.6% were teacher interactions with individual pupils; that on average teachers only spent 12% of lesson time interacting with the whole class.

The Junior School Project

In their study, Mortimore et al. (1988) put their research boundaries around schools, making the educational effectiveness at both school and class levels their research focus. The core research questions fell on whether there were differences of effectiveness between schools as well as between classes and, if so, what factors contributed to such differences. Thus, this study examined both school effects and teacher effects on pupil learning progress.

The study followed 2,000 pupils (aged 7 to 11) from 50 schools through four school years. Though not overtly claiming to do so, the researchers seemed to have employed the descriptive-correlational-experimental loop that the American researchers, Rosenshine and Furst (1973), recommended in their work. They, first of all, entered the schools and classes and described what the situations actually were, with initial variables recorded, then built connections between variables and pupil learning gains over time and finally identified the effective factors that significantly contributed to performance progress.

Overall, the effective factors that the researchers found at the end of their study were located at two levels: the school level and the classroom level. Each level was then sub-divided into two categories: '*given*' variables and *policy* variables. Thus, the effective factors were finally clustered into four domains: the school-level – 'given' variables, the school-level – policy variables, the class-level – 'given' variables, and the class-level – policy variables. The educational effectiveness at both school levels and classroom (teacher) levels were mutually supported and interwoven with each other.

The study concluded its findings with twelve key effective factors which consisted of both school-related and teacher-related factors:

- Purposeful leadership of the staff by the headteacher
- The involvement of teachers
- Consistency among teachers
- Structured sessions
- Intellectually challenging teaching
- The work-centred environment
- Limited focus within sessions
- Maximum communication between teachers and pupils
- Record keeping
- Parental involvement
- Positive climate

(Mortimore et al., 1988, p. 250)

Among these, the factors in relation to teacher behaviours were: (1) *structured sessions*, (2) *intellectually challenging teaching,* (3) *work-centred environment*, (4) *limited focus within sessions*, (5) *maximum communication* and (6) *record keeping.*

The Young Children at School in the Inner City Study

In order to investigate factors in relation to pupil learning outcomes and gains at early school years (Reception to Year 2), Tizard et al. (1988) conducted a three-year project in 33 infant schools of the Inner London Education Authority (ILEA). Four types of factors have been focused on including (1) background factors about children and their families, (2) parental involvement in home-based learning and home-school contact, (3) school and teacher related factors and (4) children's perceptions of their learning. Data were collected from pupils, their parents, their teachers and their headteachers. Schools were selected based on the criteria that each school should have pupils from multiple ethnic groups including children of

Afro-Caribbean origins and white British. Research methods included interviews, observations and standardised tests. Observations were done at the pupil level. Within each class, two girls (one Black and one White) and two boys (ibid.) were chosen as target pupils to be observed for one school day per year.

The study found curriculum coverage, teacher expectations and pleasure to teach had positive impacts on pupil learning gains in mathematics, but there were no connections between pupil progress and home variables. Moreover, it was also found that curriculum coverage and teacher expectations were independently correlated with pupil learning gains to which teacher pleasure to teach was not independently related. It was concluded that factors at the school and classroom levels explained considerably the variation of learning progress among pupils; that home and parental factors counted much less influence on pupil learning gains.

This study, again, proved the considerable effects of teaching on children's learning and also shed light on the impacts of ethnicity and gender characteristics on children's learning. Nonetheless, the highly selective nature of its sampling method limited the generalisability and reliability of its findings.

The Primary, Assessment, Curriculum and Experience (PACE) project

Following the new implementation of the 1988 Education Reform Act, which aimed at raising pupil attainment standards, Pollard et al. (1994) conducted the

PACE project. Through this project, the researchers investigated the impact of this educational policy on teachers' perspectives and their teaching practices and the subsequently indirect impact from teachers on the learning performance of Key Stage 1 pupils aged 5-7. Given the focus of this study was on the change that the newly implemented curriculum and the assessment, SATs, made to teaching and learning in schools, no sufficient evidence found could demonstrate the correlation between teacher behaviours and pupil achievement as normally seen in the process-product research on teaching effectiveness, except evidence on one significant correlation between whole-class interaction and pupil learning outcomes.

The research methods, however, were rich in detail and worthy of notice and review. The data collection methods included two rounds of both interviews and questionnaires to teachers, and two rounds of classroom and assessment observations. The main sample of the study was 48 schools chosen with a balance between various social-economic factors, urban/rural areas and geographical locations in different regions of England. The main sample provided data for advance questionnaires and the first round of interviews. The sub-sample consisted of nine out of the 48 schools from the main sample and contributed to the data for both systematic classroom and assessment observations.

The Effective Teachers of Numeracy (ETN) project

With data collected from 90 teachers and 2,000 pupils in eleven English primary schools deemed effective in maths teaching, Askew et al. (1997) sought to identify factors relating to effective maths teaching and possible ways for spreading such positive effects to wider contexts. Through the assessment of learning gains over two consecutive English school terms – Autumn and Spring terms, the study identified highly effective teachers. Data included a questionnaire with 90 teachers, lesson observations (84 lessons) in 33 teachers' classrooms, interviews with six headteachers, 54 interviews with 18 teachers and 30 validation interviews with 15 teachers.

The study had a particular focus on the relationships between teacher beliefs as well as teacher PCK and their actual teaching in the maths class. As a result, three orientations of maths teaching were identified: connectionist orientation, transformation orientation and discovery orientation. A major finding was that highly effective teachers tended to be connectionists who were able to interconnect different parts of mathematical knowledge and link the current status of children's learning and understanding to their subsequent teaching plans and practices.

Since teacher participants were from schools effective in maths teaching, the sample was not representative of the whole teacher population in England. Nevertheless, the study and its findings presented a picture of effective maths teaching in English schools in the mid-1990s and shed light on issues as to how and why certain kinds of maths teaching might have worked.

The ETN project was also one of the few studies that had employed approaches from both the maths education research community (one that traditionally heavily favours the internal qualitative factors) and the teaching effectiveness research community (one that traditionally focuses too much on the external quantitative factors). The study marked a meaningful step forward for both fields.

Mathematics Enhancement Project Primary (MEPP)

Mathematics Enhancement Project Primary (MEPP) started in 1998 and lasted for over two years until 2001. It aimed to transform the traditional British way of mathematics teaching in primary schools from teaching with a high proportion of individual work towards whole-class interactive teaching (Muijs & Reynolds, 2003).

The results of this project reflected and compared the effectiveness of mathematics education, both within the boundaries of schools and across two geographical regions of the UK. The data consisted of observations of 138 teachers and standardised tests to 4,813 pupils in 36 primary schools in England and Wales. The observation was conducted using the *Mathematics Enhancing Classroom Observation Recording System* which measured 57 teacher behaviours against a Likert scale ranging from 1 to 5. Pre- and post-tests were done at the beginning and the end of each academic year to provide accurate measurement of pupil learning outcomes and gains over time.

The concentration of the analysis was on the relationship between teacher behaviours and pupil learning outcomes/gains. By setting variables in four models, multilevel modelling was applied in two domains – learning outcomes and learning gains – to analyse variables at three different levels, i.e., pupil, class and school levels, connecting these variances with both pupil test scores at certain points in time and their learning gains over a period of time.

The final results revealed that 24% of pupil achievement differences at the class level could be explained by teacher behaviours and over 50% of differences in learning gains over two years at the class level could be attributed to teacher behaviours.

2.5.3 Effective teacher behaviours identified

Applying most of the lesson time to whole-class interaction

Among the factors that British research has identified as evidence for effective teaching, whole-class interaction is the commonest, either internally in this country or externally compared with American research findings.

The ORACLE study (Galton & Simon, 1980; Galton et al., 1980) found that pupils were left working on their own in two thirds of the class time and that teachers spent 71.6% of their limited interaction time interacting with individuals about routine things, most of which were to get children involved in their tasks.

The way in which university tutors interacted with individual students through personal tutorials was found to be extremely popular in the primary classrooms where around 29 other pupils could not be simultaneously engaged into the one-to-one communication with their teacher. Such individualised interaction was time- and energy-exhausting for the teacher and unfair for the other pupils.

The amount of time, the audience and content of teacher attention were also observed and recorded systematically, which made possible the quantitative analysis of teacher behaviours in this manner (ibid.). The researchers found that teachers' attention was nearly equally distributed to individual pupils across the class whether they were high, medium or low achievers. The only difference was that teachers gave low achievers slightly more attention. In addition, there was no overt difference of attention distribution from teachers to the two different genders, though boys were found to be receiving slightly more contact than girls. There was, of course, one factor, class size, that was negatively correlated with the average amount of individual attention. Overall, teachers allocated 47.5% of lesson time paying attention to individual pupils, and 39.3% out of the 47.5% was exactly task-oriented. The result showed that teachers were using about half of the lesson time in meeting individuals' needs, either through one-to-one communication or interaction in small groups instead of interacting with the whole class.

The typical phenomenon of teachers' intended attentions to individual pupils in many British classrooms coincided with the progressive teaching approach recommended by the Plowden report (Central Advisory Council for Education, 1967) which stressed child-centred education and the individualisation in classrooms and followed Piaget's (1962) theory on the development of children's cognition. Nonetheless, as the researchers (Galton et al., 1980) argued, the individualisation recommended by the Plowden report was not realistic for the classroom where around thirty children's needs were waiting to be met and teachers could hardly have sufficient time and energy to cope with each pupil individually. This argument was consistent with Muijs and Reynolds' (2003) findings in MEPP that pupil learning outcomes and gains were both positively correlated with the amount of time for active whole-classroom teaching in lessons.

In the ORACLE study (Galton et al., 1980, Ch.6), teachers were categorised according to their teaching styles as four main types: (1) individual monitors, (2) whole-class enquirers, (3) group instructors and (4) style changers. The last type was further classified as three sub-types: (4a) infrequent changers, (4b) rotating changers and (4c) habitual changers. The whole-class inquirers were found to only be focusing on the subject content they planned to teach, mostly teaching to the whole class, and therefore maximising their interactions with everyone. Follow-up analysis of the study (Croll, 1996) proved a positive correlation between whole-class interactions and pupil learning outcomes, which was consistent with evidence from other British studies (Muijs & Reynolds, 2003; Pollard et al., 1994) and early American research evidence (Evertson et al., 1980; Good et al., 1983).

A similar finding from the Junior School Project (Mortimore et al., 1988) was that, on average, teachers spent 67% of lesson time in interacting with individual pupils, leaving just 23% of lesson time for whole-class interaction. In mathematics lessons, 13% of lesson time was spent in teacher questioning, whereas just a quarter of the questioning time, i.e., only about 3% of the lesson time, was applied at the whole-class level. Nevertheless, teaching and interacting was not necessarily positively correlated with progress in learning outcomes unless such activities were efficiently and thoroughly woven into the whole range of the class (ibid.).

Strong knowledge of subject matter, pedagogical content and learners

The ETN project found that highly effective teachers had deeper understandings of mathematics, pedagogic content and pupils. They did not necessarily demonstrate a large amount of mathematical knowledge – " more is not necessarily better", but what they did have were deeper understandings of the connectedness of the subject knowledge (Askew et al., 1997, p.98). They utilised assessment to systematically diagnose children's knowledge and then design their next-step teaching accordingly, which helped improve learning on a regular basis. Less effective teachers tended to merely care about whether the results were correct or wrong, and when children did not get the knowledge, they simply re-taught the stuff and asked for more practice.

More time on task

Evidence from the ORACLE study showed a positive correlation between whole-class interaction and time on task (Croll, 1996). In the ORACLE study, researchers documented the proportion of time that teachers and pupils spent at class in the teacher records and pupil records, which then provided raw data for in-depth analyses. The subsequent analysis on timing of teacher-pupil interaction revealed the teacher participants spent an average of 71.6% of interaction time interacting with individual pupils, whereas only within 58.1% of lesson time were pupils actively engaged into the task. In the *One in Five* study, which focused on the correlations between interaction types and academic engagement, Croll and Moses (1985) collected and analysed data from 32 Year-4 classrooms finding a considerably positive correlation between whole-class interaction and time on task and a moderate negative correlation between individual interactions and time on task. One third of the amount of lesson time was found to be allocated to curricular tasks in the classes where teachers applied the whole-class interaction approach, whereas no more than 45% of lesson time was found to be allocated to academic tasks in classes where teachers implemented few whole-class interactions. Similarly in the PACE study, Pollard et al. (1994, p. 182) also found that, "... classrooms where high levels of class interaction were recorded are also likely to have been coded with high levels of pupil task engagement ..."

Managing the classroom seating and organisation appropriately

As Croll (1996) argued, whole class interaction did not mean teachers should give lectures in a monologue and top-down manner with the pupils seated in the conventional formats. Conversely, such interaction must be two-way dialogues between teachers and pupils who were seated in a way well suited for communication and more accessible to their teachers. The horseshoe seating was shown to be an effective classroom layout, which not only supported all the three types of interactions with individuals, groups and the whole class, but also made it easy for teachers to maintain a balance between various types of classroom interactions (Galton et al., 1999; MacNamara & Waugh, 1993).

Appropriate teacher expectations

Teacher expectations have long been identified as an effective factor that positively correlates with pupil learning outcomes/gains. British research provides strong evidence for such correlations. For example, Tizard et al. (1988) noted in their study:

> … the association between teacher expectations and [children's] progress in maths during reception [is] 0.4. This means that the difference in maths progress from the top of the teacher expectation scale—that is, those rated above average—and the bottom of the teacher expectation scale—that is, those rated below average is 0.4 S.D., a statistically significant difference... expectations are significantly related to progress in each year...
>
> (Tizard et al., 1988, pp. 138-139)

A positive correlation between lower expectation and decreased pupil academic performance was also discovered in the study by Mortimore et al. (1988), as signals of lower expectations were transmitted to pupils through teachers' special attention based on pupils' SES, race and gender.

Limited focus in lessons and sufficient curricular coverage over time

In the literature, it is evident that teachers who limited their focus to one or two curricular areas in a lesson were more effective than their peers. In addition, the amount of curriculum content that teachers covered at class over time was shown to positively correlate with pupil learning outcomes.

The Junior School project (Mortimore et al., 1988) found a disparity between what teachers reported to have done and what they were observed having actually done. Teachers believed that their teaching was more focused on a single curricular area, while the researchers observed and found that those mathematics teachers merely spent 18% of lesson time on the content of mathematics. Making comparisons across schools, Mortimore et al. found that more effective teachers oriented their lessons at no more than two curricular areas at the same time. The progress at the school level was consistent with the progress at the class level in Tizard et al.'s (1988) study, and a larger curriculum coverage was found along with higher academic expectations from teachers as the factors that contributed to the increase of pupil test scores. The researchers also found that teacher expectations tended to be transmitted to pupils through the curriculum content that the teacher assigned them to work on at class and/or at home.

2.5.4 **Conclusion**

The review of early British research on teaching effectiveness helps reveal: (1) what methodologies and designs British researchers had generally employed and how well and (2) what conclusions they had explicitly come up with. The methodology commonly used was building connections between observed teacher behaviours and pupil achievement/attainment, and the effective factors identified tended to work collaboratively rather than separately. During the reviewing process, the gaps in those two domains also naturally emerged. The following paragraphs will serve to explain them in detail and, through the explanation, seek indications for the design of this study.

The methodologies and research designs of British TER studies – particularly those of the four major studies – formulated a picture with a wide scope. Building on often large-sized samples, TER studies in the UK inherited quantitative characteristics from TER in the US and continued identifying key teacher factors that had considerable impacts on pupils' learning outcomes/gains. Those studies added extra values to the body of knowledge on what works and what does not in classrooms and schools. Nonetheless, there is still a lack of detailed explanations beneath the quantitative results. Moreover, given that there was a dearth of a simultaneous focus on the activities of all characters, particularly pupils, in the classroom, the observational techniques applied in the past could not present the full complex picture of the dynamic interactions in a class, hence missing the mark in answering research questions thoroughly, and would fail to compete against today's observational demands and strategies.

Matching the research findings from British and American studies, there are certain effective factors commonly identified. However, there are also many other factors that were less frequently identified and discussed by British researchers but were either frequently noticed by TER researchers from other countries or recommended by educational theorists and researchers utilising methods other than the product-process approach.

The next section will mainly focus on the current state of TER studies conducted in various countries.

2.6 National studies on teaching effectiveness

2.6.1 Introduction

International studies of pupil performance revealed the disparity of the academic outcomes between educational systems across different countries. In parallel with the international investigations on teaching and schooling among countries and systems, there has been an on-going trend within a number of countries to examine factors that decide the effectiveness of their national education. Most of those national studies checked the educational effectiveness at either the school level or the classroom level.

In the theoretical domain, the two-level effectiveness categories have been combined recently with other aspects of effectiveness research and clustered as a holistic field – educational effectiveness research (EER). In EER, educational outcomes are increasingly seen as being formulated by various types of factors working collectively at multiple levels - the pupil, the classroom, the school, and the system levels (Creemers & Kyriakides, 2006; 2008).

In this section, firstly the history of educational effectiveness research will be reviewed, merged with the development of school effectiveness research and teaching effectiveness research. Secondly, evidence from previous studies will be presented to explain the difference between teacher effects and school effects. Thirdly, findings from a number of national studies will be presented and discussed. Finally, a short conclusion will reflect on the key themes on the basis of what has been reviewed in this section and identify the research gaps that the current study intended to fill in.

2.6.2 Educational Effectiveness Research

One of the earliest areas in EER is the School Effectiveness Research (SER) which historically had its 'first' voice in responding to Coleman et al.'s (1966) assertion that schools made no difference and the subsequent long-term defence by many other researchers regarding the actual effects of schools and the functions of the educational system, either outside or inside the schools. As follows, the developmental stages of EER will firstly be traced, and then there will be a brief overview of the main aspects that EER researchers aim to measure.

History of EER

Based on many previous reviews of the history of EER (e.g., Creemers et al., 2010; Mortimore, 1991; Reynolds, 2010; Sammons, 1999; Teddlie & Reynolds, 2000), the development of this field has been through five phases (Reynolds et al., 2011).

The first phase of EER started as a reaction to the Coleman Report (Coleman et al., 1966), as mentioned before. Then followed the empirical studies conducted by Edmonds (1979), Rutter et al. (1979) and Smith and Tomlinson (1989) and small-scale case studies carried out by Weber (1971) and Reynolds (1976).

The second phase occurred from the mid-1980s when multilevel methodologies (Goldstein, 2003) and other studies applying such methodologies emerged, and consistent educational effects in certain areas in the long term proved the scientific properties of school effectiveness.

The third phase came in the early to mid-1990s as numerous researchers attempted to explore the reasons that caused schools to perform differently in educating children. In the US, there was the Louisiana School Effectiveness study conducted by Teddlie and Stringfield (1993), and in the UK, Sammons et al. (1997) researched into the effects of departments on pupil performance and the effectiveness of schools. Such actions were also realised through a number of prominent reviews of the field (Creemers, 1994; Reynolds et al., 1996; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000; Teddlie & Stringfield, 1993).

The fourth phase started in the mid-1990s and went through a decade with its influence and development remaining today. The remarkable theme in this phase is the internationalisation of the EER field, and the close connections and collaborations of school effectiveness (SE) researchers with school improvement (SI) practitioners. Research on educational effectiveness has travelled from a few initial countries to a wide range of countries on the planet, such as the Dutch study by Bosker and van der Velden (1989), the French study by Grisay (1996), the Australian study by Hill and Rowe (1996), Flanders' studies by Van Damme et al. (2006), De Fraine et al. (2007) and Verachtert et al. (2009). The wider geographical development of EER also formed the trend of internationally joint research in the field. Within the field, SE and SI researchers are becoming more comfortable with each other's methodologies in that SE's quantitative perspective gets in line with SI's qualitative vision and the two collaboratively investigate educational issues through mixed methods.

The fifth phase started at the end of 2000s and is currently still rapidly developing. The focus of EER in this phase mainly addressed the dynamic, rather than static, characteristics of relationships between all the effective components in EER, and the impact of the interactions between these components on educational outcomes (Creemers & Kyriakides, 2008).

The focus of the measurement in EER

Although the prevalent criterion in EER is still pupil academic performance, there are also gradually more and more researchers who pay attention to other aspects of educational outcomes, such as pupil well-being (Konu et al., 2002; Van Landeghem et al., 2002) and progress drive (Van de gaer et al., 2009; Van der Werf et al., 2008). In addition, pupils' learning outcomes are gradually becoming more researched over a longer time span, which thus helps judge how long the values of educational effectiveness can last and to what degree such effects can follow the children through their later schooling or their life after leaving schools.

Children are organised in classrooms which are netted in schools and then systems. EER researchers have been concerned with the question as to which educational level makes a bigger difference. Thus, effect sizes of different levels have been compared either through empirical investigation or theoretical meta-analysis. The most scrutinised pair is schools versus teachers. Next section, we will look at the effect size of teachers as apposed to that of schools on pupil learning outcomes, as suggested by the literature.

2.6.3 Teacher effects vs school effects

The size of teaching effects has consistently been found much higher than that of school or system effects. Many studies conducted within various countries have all come to the same finding that individual teachers make a significantly stronger impact on pupil performance than schools do.

Table 2.3 Teacher and school effects in eight countries

Country	*Class/Teacher Effects (%)*	*School Effects (%)*
Canada	17	9
Finland	45	0
France	16	6
Israel	21	8
New Zealand	42	0
Scotland	31	5
Sweden	45	0
USA	45	9

Source: Scheerens et al. (1989, p. 794).

Through a secondary analysis of the results from the Second International Mathematics Study (SIMS), Scheerens et al. (1989) found the degree of educational impact on pupil achievement varied considerably from the classroom level to the school level with evidence supported by data from eight countries (see Table 2.3).

There are also similar findings from a number of Australian studies. Investigating the teaching and schooling of both primary and secondary literacy and numeracy, the Victorian Quality Schools project (Hill & Rowe, 1995) examined the educational effects on 13,700 pupils from 99 primary and secondary schools. The comparisons of effects at two levels on pupils' learning outcomes are presented in Table 2.4 from which it can be clearly calculated that, from primary to secondary and from literacy to numeracy, teacher effects are approximately between 5 and 10 times bigger than school effects.

Table 2.4 Teacher and school effects for Victorian schools

Subject	*Phase*	*Class/Teacher Effects (%)*	*School Effects (%)*
Literacy	Primary	45.4	8.6
	Secondary	37.8	7.4
Numeracy	Primary	54.7	4.1
	Secondary	52.7	8.4

Source: Hill and Rowe (1995, p.20).

A similar conclusion has also been drawn by Cuttance (1988):

> Recent research on the impact of schools on pupil learning leads to the conclusion that 8-15% of the variation in pupil learning outcomes lies between schools with a further amount of up to 55% of the variation in individual learning outcomes between classrooms within schools. In total, approximately 60% of the variation in the performance of pupils lies either between schools or between classrooms, with the remaining 40% being due to either variation associated with pupils themselves or to random influences.
>
> (Cuttance, 1998, pp. 1158-1159)

In the Canadian study – 1996 Elementary School Climate Study (Willms, 2000), bigger differences in pupil performance were found between classes than between schools and between districts (Table 2.5).

Table 2.5 Variation among school districts, schools and classes

Measured Outcomes	*Between Districts*	*Between Schools*	*Among Pupils Within Classes*
Reading	0.3	0.8	98.9
Writing	1.0	3.4	95.5
Mathematics	1.8	4.7	93.5
Science	0.4	3.8	95.8
Self-esteem	0.1	3.0	95.8
Sense of belonging	0.3	1.0	98.7
General well-being	0.4	1.6	98.1
General health	0.8	0.0	99.2

Source: Willms (2000, p. 241).

A synthesising review of 500,000 studies by Hattie (2003) from New Zealand again indicated the essential role of teachers as a main source of pupil learning improvement. It is clear that, in Table 2.6, pupils (i.e., their ability level, previous performance level, etc.) and teachers at the classroom level explain the majority of performance variance. Internally, pupils' in-born ability is not easy to change and their previous performance might be formulated either by their in-born intelligence or the education they have previously experienced. Externally, it is clear that teachers are the crucial factor that contributes the most to pupil performance.

Table 2.6 Percentages of achievement variance

Causes	*Pupils*	*Teachers*	*Schools/Principles*	*Peer*	*Home*
Percentages	50%	30%	5-10%	5-10%	5-10%

Source: Hattie (2003, p. 3)

Another meta-analysis done by Luyten (2003) reviewed a number of empirical studies, comparing the size of school effects to that of teacher effects across subjects, parallel classes and grades, and found:

- ✧ teacher effects generally outweigh school effects across subjects, parallel classes, and grades, but are smaller than school effects when compared across parallel secondary-school classes;
- ✧ teacher effects across the two subjects, mathematics and language, are larger in primary than in secondary, whereas school effects are smaller in primary than in secondary;
- ✧ mathematics teachers' effects appear to be larger than language teachers' at the primary phase, with the case reversed at the secondary phase.

Although the overall results indicate a stronger influence from the classroom level than from the school level on learning outcomes, Luyten further mentioned the inevitability of the dominance of teacher effects, and discussed the possibility of balancing the two types of educational effects by reducing class differences and transforming the class-level strengths to the school level. If the prediction could be proven to be true, it is definitely inspiring, since successfully educating a whole school of pupils will make greater impact than successfully educating pupils in only one class. However, there is a lack of sufficient empirical evidence to support Luyten's (2003) prediction. As Hill and Rowe (1996) argued, it is mainly teachers that make schools work differently, although schools do make a difference. The reviews done by Reynolds and Teddlie (2000) and Scheerens and Bosker (1997) came up with the same finding that teacher effects are much larger than school effects.

2.6.4 Effective teaching factors in national studies

The aforementioned Hattie's (2003) review on half a million studies comes up with 5 dimensions and 16 attributes of effective factors that expert teachers generally have. The review suggests that effective teachers are able to represent the subjects optimally, create appropriate environment for best learning to happen, diagnose and tackle learning difficulties, enthuse about pupils and teaching and learning and engage children towards effective learning.

Involving 50 Australian schools, Ingvarson et al.'s study (2004) focused on the correlations between teacher knowledge and pupil learning outcomes. Data were collected through two tests, questionnaires with pupils, teachers, headteachers and Maths departmental leaders and case studies. The study found that teachers' educational background and the professional quality of the Maths department were strong predictors of learning gains and that teaching methods predicted pupil perceptions of Maths learning.

Starting from the theories of learning, Munro (1999) followed 32 Melbourne teachers through the process of a professional programme, called Facilitating Effective Learning and Teaching, and identified the changes of teaching behaviours coincided the learning theories embedded in this programme. Such types of teaching behaviours were also in line with the social-constructivist model of learning (Brown, 1994; Voight, 1994). The aim of the programme was to promote teachers' reflection on pupils' learning in class under a framework containing nine components of learning so that teachers could enhance their teaching behaviours in nine areas:

1) Facilitating the efficient allocation of pupil attention while learning;
2) Facilitating the learning of positive attitudes towards content area learning;
3) Providing pupils with a range of options for learning an idea particularly when they find learning difficult;
4) Providing pupils with a range of options for displaying what they know or have learnt;
5) Helping pupils to organize themselves more efficiently as learners in their subjects;

6) Using learner-oriented behaviour management strategies such as using on-task redirection strategies;
7) Providing pupils with the opportunity to learn more how they learn;
8) Providing pupils with the opportunity to take control of their learning in a more collaborative relationship;
9) Providing pupils with the opportunity to monitor their own learning and to see themselves progressing.

(Munro, 1999, p. 162)

Pre- and post-measurement was done in terms of changes in three domains: changes in teachers' behaviours, changes in teacher knowledge and changes in pupil learning outcomes. Teaching behaviours in the nine components were significantly improved, with an average increased number of each type of teacher behaviours at the end of programme ranging from 4.6 per lesson to 8.4 per lesson and, two terms later, ranging from 4.7 per lesson to 9.8 per lesson. Clearly, teachers significantly changed their behaviours in the classroom and increased frequencies of learning-oriented teaching behaviours after participating in the programme. Additionally, such changes increased as time went on. Similar changes were found in teacher knowledge about learning, and pupil performance across subjects was positively increased by 127 scores, of which 70% was actually attributed to the improvement of the *quality* of learning. Thus, the study found a positive correlation between the change of teaching behaviours and the change of pupil learning outcomes. However, there were also defects in the report of the study: (a) the measurement of pupil learning gains was not clearly presented; (b) the criteria of learning assessment were not explained in detail.

In a Dutch study conducted by Luyten and de Jong (1998), teacher effects and school effects were examined in the subject of mathematics in parallel classes within and across 20 secondary schools. Through structured interviews and common tests in parallel classes, pupil achievement was compared at the pupil level, classroom/teacher level and school level. Apparent differences were found within classes, while pupils from parallel classes performed at a similar level, i.e. there were little differences of learning outcomes between parallel classes at the same grade – the first year in secondary schools. Given the convenience of comparing pupils' achievement across parallel classes, teachers of those classes were able to unify their instructional content and teaching goals, which overall might explain the less variation of pupils' performance across parallel classes than across subjects and grades. Controversially, measured against a five-aspect model, which contained structured instruction, clarity, pupil enjoyment, teaching content, and homework quantity per week, mathematics teachers were found significantly different from each other within schools in all five aspects, particularly in terms of structured instruction and homework quantity per week.

Another Dutch study done by den Brok et al. (2004) looked into the correlations between teachers' interpersonal behaviours as perceived by pupils and pupils' learning outcomes. Findings were based on data derived from 45 physics teachers, 32 English teachers and the corresponding Year-3 pupils of these teachers. Two components of pupils' perceptions of teacher interpersonal behaviours, influence and proximity, were shown to be responsible for more than half of the learning-outcome differences. Overall, Physics teachers' interpersonal behaviours had stronger impact on pupils' affective outcomes but weaker impact on pupils' academic outcomes than did English teachers'. This is consistent with some other studies' findings in that teacher effects vary between different domains of outcomes and across various subjects (Luyten, 1998; Reynolds, 1995).

In 2012, Opdennakker, Maulana and den Brok (2012) reported their latest study focused on the same types of teacher behaviours. The research foci were on the developmental trends of teacher-pupil relationship and pupil academic motivation over the first year of secondary schools. Data were collected five times from 566 pupils of 20 classes taught by 10 mathematics teachers and 10 English teachers in 3 secondary schools in the Netherlands and analysed through multilevel growth curve modelling. The study found that the relationship between the two focused variables was more related to influence than proximity and that pupils' controlled academic motivation increased throughout the school year, whereas their autonomous motivation decreased as time went on.

In Germany, Einsiedler and Treinies (1997) conducted a study which combined both experimental and observational data. Twenty-one grade-four classes were randomly and averagely divided into three groups of classes which were taught by teachers using three different representational methods, i.e., hierarchical knowledge structuring, network like knowledge structuring and traditional knowledge structuring (in the control group). In the experimental section of this study, researchers found that pupils above average benefited more in the treated groups, but pupils below average progressed more in the controlled group. However, overall, it was found that class effects were stronger than the effects of teaching methods, i.e., the representational methods. Utilising Flanders' method (1970), trained observers recorded the frequencies of various types of verbal cognitive interactions between teachers and pupils in the 21 classes. Consequently, the observational data revealed more directions in the two treated groups than in the controlled group. In addition, teachers in treated groups tended to structure lessons with graphic representations, while teachers in the controlled group were more likely to apply cognitive structuring. Overall, the researchers stressed the importance of class contexts to the effects of school education and the need to take into account the role of teachers in formulating the contexts of classes.

In a project on stakeholders' perceptions of education quality in China, Peng et al. (2014) conducted individual and focus group interviews with over 90 stakeholders, including headteachers, teachers and high school students from 8 secondary schools across three LEAs in China. Partial findings of the study indicated that Chinese teachers were expected to play multi-functional roles, that they experienced dilemmas in meeting the demands of the curriculum reform, that the formulation of a positive learning environment in a school was expected to be a collaboration between teachers and pupils, that within Chinese schools there existed professional learning communities which aimed at improving teaching through collaborations between teachers, their colleagues and/or external experts from teacher education institutions. The study also found the disparity of educational development between western rural areas and other parts of China and the dilemma that stakeholders in the West of China faced as they strove to improve teaching and learning in their schools.

Another noteworthy TER study was conducted by Teddlie and Liu (2008) in a relatively poor province, Jilin, in Northeast China. The sample included teachers and pupils from 8 urban and 4 rural schools. Collecting data through observations of 300 lessons in five grades (year levels) across12 schools, the study found that effective teaching factors identified in the West actually worked in these Chinese schools, that more effective schools had teachers demonstrating more effective teaching than less effective schools, and that urban schools tended to have more effective teachers than rural schools. The study identified effective teaching factors, such as nurturing positive learning environment, maximising the proportion of lesson time spent teaching, managing pupil behaviours and interactive time-on-task. A set of teaching factors, including whole class interactive teaching, was believed to have guaranteed the high average of interactive time-on-task (M = 81%) – a strikingly much higher score than the combined American score (44%) in the literature. However, whether the performance differences between countries (such as China and the US) were attributed to teaching differences could only be answered by *cross-national* TER studies.

In this section, we have looked at TER studies carried out in Australia, the Netherlands, Germany and China. Due to word limit, many other studies from

Belgium (e.g., Boonen et al., 2013), Cyprus (Kyriakides et al., 2009), the Netherlands (Driessen & Sleegers, 2000; Meijnen et al., 2003) and other countries (Konstantopoulos & Sun, 2013; Palardy & Rumberger, 2008; Teodorović, 2011; van de Grift, 2013) will not be reviewed here.

2.6.5 Conclusion

The field of EER has evolved vertically through over four decades and horizontally across locations – classrooms, schools, and countries. Similar and different effective educational factors have emerged among EER studies most of which were conducted within countries. The flourishing of EER in more countries offers good opportunity for researchers in the field to carry out international studies to examine what factors work internationally and what may work better locally.

In terms of paradigm orientation of the field, there is a severe lack of rich descriptions and explanations in the outputs of EER where quantitative approaches have been playing a dominant role and qualitative evidence has rarely been collected.

Decades of educational effectiveness research consistently prove the crucial role of classroom-level factors, particularly teacher behaviours, in fostering the improvement of pupils' learning. In line with the international comparisons of educational outcomes at similar ages, there is a need for TER researchers to conduct cross-national studies on the effectiveness of classroom teaching in order to generalise internationally valid and reliable predictors for pupil performance and promote the development of and universal applicability of educational science (Reynolds et al., 2002a).

2.7 Chapter conclusion

This chapter reviewed five domains of literature in the field of educational research: (1) mathematics pedagogical theories and practices, (2) international comparisons on learning outcomes and teaching practices in school mathematics, (3) early research on teaching effectiveness in the US, (4) early research on teaching effectiveness in the UK, and (5) other national studies on teaching effectiveness.

Works on mathematics pedagogy are either theories in themselves or deeply rooted in theories. Empirical efforts are generally devoted to the learning of mathematics and teacher subject matter knowledge. Evidence about what works in the teaching of mathematics is still fragmented and somehow recondite.

International studies on learning outcomes, particularly in the subject of mathematics, revealed the educational effectiveness of a country/region in comparison with those of other countries/regions. The corresponding investigation into instructional practices among countries paid less attention to sufficiently and systematically correlating teaching with learning across countries, and hence has not come up with internationally generalisable findings.

The early research on teaching effectiveness in the US and the UK revealed a slight shift from the process-product research towards not only the correlations between teacher behaviours and pupil learning outcomes and/or gains but also deeper inquiries into the reasons lying beneath the surface of various correlations. Later research on teaching effectiveness in other countries revealed various types of research foci and enquiry approaches. Mixed-methods gradually came into the research trend to complement the quantitative methods with great details collected via qualitative methods. Overall, this adds into EER, including the TER, down-to-earth properties, bringing the area in line with the educational improvement research to make greater impact on policy making and the educational practice across countries.

To address the research questions (see section 1.6), a variety of methods were applied in collecting and analysing data, which are detailed in the following chapter.

CHAPTER 3 - METHODOLOGY

Chapter overview

- *Chapter introduction*
- Research philosophy
- Overview of the study and its design
- Data collection and analysis methods
- Ethical issues
- *Chapter conclusion*

3.1 Chapter introduction

In the last chapter, we have looked at international studies on learning outcomes and teaching practices in mathematics, reviewed the evolution of TER and EER over time and across countries and reflected upon the key findings and methodologies of the subfield – TER – and its parent field – EER. Out of the review of relevant literatures emerge the research gaps that this study intends to fill in and the three main research questions that the study aims to address. In this chapter, we will focus on the technical part of the project to explain in detail:

- the research philosophy in which the study is embedded
- the research design which defined and was defined by the study
- specific data-collection and data-analysis methods applied
- potential ethical risks and measures taken to omit such risks

3.2 Research philosophy

Living with both physical and spiritual features and in a combination of both objective and subjective matters, human beings have different epistemological views in distinguishing knowledge and non-knowledge and various ontological views in defining "what exists, what is the nature of the world, [and] what is reality" (Usher, 1996, p. 11). Around such differences in understanding the fundamental issues of the world line up the long-lasted paradigm war between research communities that hold different beliefs. Arguments are mainly around "the form and nature of reality", "the relationship between the knower and what can be known", and the approaches (or methods) that "the inquirer go about finding out what can be known" (Punch, 2009, p. 16). As follows, the debate between two typical paradigms in academia – positivism and interpretivism – is described, and then an overview of the rise of the third paradigm – the mixed-method approach – is presented.

3.2.1 The war between paradigms: positivism vs. interpretivism

Kuhn defines the paradigm as "the entire constellation of beliefs, values, techniques shared by members of a given scientific community" (Kuhn, 1970, p. 75) and "an exemplar or exemplary way of working that functions as a model for what and how to do research, what problems to focus on and work on" (as cited in Usher, 1996, p. 15). The seemingly close relationship between research methods and the philosophical values behind them raises a question about which most researchers have to think before moving on to technical details – that is to clarify the researcher's philosophical position and the ways it shapes what is to be researched and how it is to be researched.

The main paradigm debates are between two streams of thoughts, i.e. positivism and interpretivism (or constructivism, or naturalism, etc. according to different terminological definitions) (Punch, 2009, p. 18). The former holds an objective view of the social world that can be observed, described, investigated, experimented and so on in a similar way that natural scientists do with physical phenomena (Oldroyd, 1986). Conversely, the latter, regardless of the different labels they have been given, generally sees the world as multiple realities subjectively viewed or constructed by human beings either individually or collectively (Cohen et al., 2007, Ch.1). Whereas positivism is often associated with quantitative (QUAN) methods, interpretivism or constructivism is generally connected with qualitative (QUAL) methods (Punch, 2009). It was said that the whole 20th century had been dominated by the "polarization of qualitative and quantitative research" (Hartas, 2010, p. 26). Alongside the dominant climate over the paradigm separation, other researchers argue against the monomethod bias that either approach might place on research findings and point out that a better solution is to choose an optimal mixture of multiple methods that are most helpful in addressing the research question(s) of a study (e.g., Bryman, 2006; Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998). The following section is concerned with this *third* paradigm and the philosophy that underpins the paradigm.

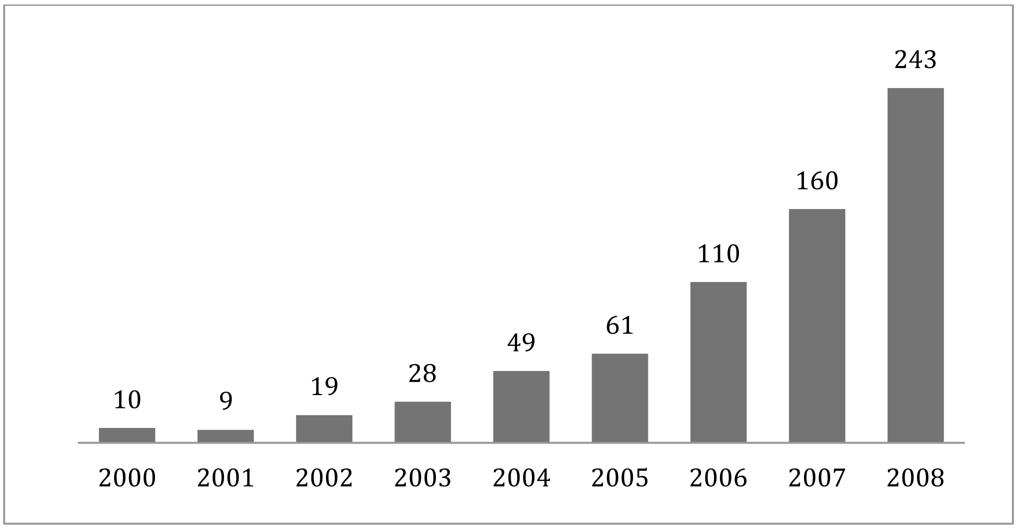
3.2.2 The rise of the third paradigm: mixed-method research

Rather than stick to either paradigm, the present trend in most social science research fields is to make use of strengths of two paradigms by mixing various methods from both paradigms. Thus rose the third paradigm, mixed-method research (MMR) (Tashakkori & Teddlie, 2010). The seemingly solid distinction between qualitative research and quantitative research is regarded as problematic in that "research in itself can be neither qualitative nor quantitative" as "only data can properly be said to be qualitative or quantitative" (Biesta, 2010, p. 98).

This paradigm has an orientation of pragmatism in that it "rejects the 'either-or' decision points associated with the paradigm wars" and prioritises the role of research questions "over considerations of either method or paradigm" (Tashakkori & Teddlie, 1998, p. 167). The use of multiple methods offers the researcher opportunities to "attack a research problem with an arsenal of methods that have nonoverlapping weaknesses in addition to their complementary strengths" (Brewer & Hunter, 1989, p. 17). It uses both the deductive logic of the QUAN and the inductive logic of the QUAL in "the research cycle"; it appreciates that the researcher's points of view might be a dynamic combination of "both objective and subjective" in the research process and the degree of objectivity or subjectivity may vary over the course of enquiry (Tashakkori & Teddlie, 1998, pp. 24-25). It is argued that researchers in the social and behaviourial sciences should put the importance of research questions over that of paradigms. Tashakkori and Teddlie (1998) put it nicely:

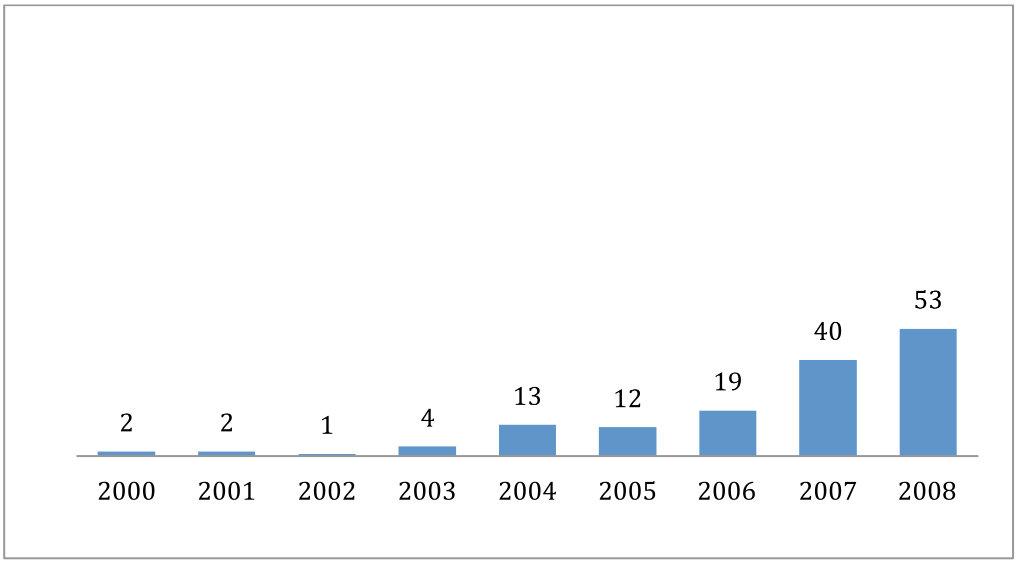
> For most researchers committed to the thorough study of a research problem, method is secondary to the research question itself, and the underlying worldview hardly enters the picture, except in the most abstract sense. (p.21)

With these practical features of this paradigm, the new millennium has seen a dramatic increase of the number of MMR studies. For instance, Ivankova and Kawamura's (2010) meta-review on research articles between the years of 2000 and 2008 found a total of 689 mixed-method empirical studies over time and a significant increase of the amount of MMR work by year both comprehensively (Figure 3.1) and specifically in education (Figure 3.2).



Source: Ivankova and Kawamura (2010, p. 591)

Figure 3.1 Mixed methods empirical studies by year



Source: ibid. (p. 593)

Figure 3.2 Mixed methods empirical studies in education by year

More specifically, like many other social and behavioural sciences, the field of EER has also developed a strong quantitative trait since its childhood (Reynolds et al., 2011). In the trend towards more application of MMR in other educational research areas, EER researchers are also calling for the adoption of the QUAL methods to supplement QUAN approaches within one study, so as to triangulate findings of various methods, to make research findings more sensible and accessible to more audiences, particularly, to practitioners and policy makers and to improve educational practices at the levels of classrooms, schools and systems (Harris et al., 2013).

Considering the research questions of the EMT project, the researcher chose a list of methods and applied them in a way that was deemed to serve the research purposes better. Next, the actual design of the study is explained in detail.

3.3 Overview of the study and its design

This short section offers a bird's eye view over the study mainly with regard to its research methods, theoretical framework and sampling method and procedures.

3.3.1 From questions to methods

As presented in Chapter 1, the aims of the study are, on the one hand, to test the correlations between teaching behaviours and pupil learning outcomes across England and China, and on the other hand, to listen to and interpret multiple voices regarding the quality and effectiveness of mathematics teaching within and across these two countries. In doing so, the MMR approach was chosen to answer the aforementioned research questions. QUAN methods were mainly applied to answer Q1, whereas QUAL methods were specifically chosen for Q2 and for the analysing process towards a meaningful conclusion of Q3.

Given that the research aim is to investigate the effectiveness of mathematics *teaching*, the classroom (teacher) level factors were put at the core of the research focus. Research methods applied include video-recorded classroom observations (both structured and unstructured), video-stimulated interviews with teachers, video-stimulated focus groups with teachers, two questionnaires for teachers and pupils respectively and two standardised mathematics tests for pupils at two points in time. Next, the theoretical framework of the EMT project is shown and illustrated.

3.3.2 The theoretical framework

Based on the research questions, methods applied and interconnections between research questions and methods and between results and findings of the project, a theoretical framework is constructed (Figure 3.3). The framework links research questions to research methods, with the left hand side of the framework standing for the first question, the right hand side the second one and, in between, question three.

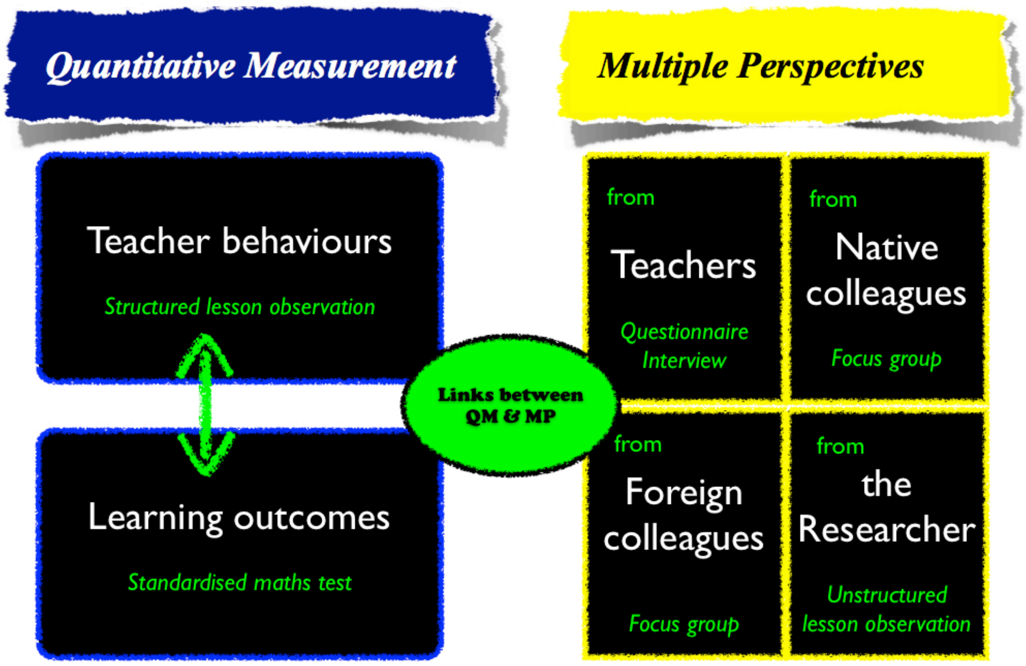


Figure 3.3 The EMT theoretical framework

The research process also proves that the theoretical framework is more than just a framework. It is also a research map that guides every step of the study. Holistically, it is also a representation of reality where numbers as well as voices, quantity as well as quality and external behaviours as well as internal perceptions inter-exist. In a sense, the framework not only structured the data collection and data analysis processes, but also illustrated the interconnections between various research methods and different components of data.

To summarise, the framework plays multiple roles in the study, in that it explains the philosophy on which the study stands, links the research questions to the research methods, and explicitly guides the preparation, implementation, analyses of and subsequent conclusions of the project.

With this research map handy, the next thing to do is to find participants. The focus of the following section is on introducing the sampling method, criteria and process of the EMT project.

3.3.3 The sampling method and procedures

Before dipping into the sampling details of the project, we will first of all listen to and hear what the methodological literature has to say regarding sampling. Sampling methods are generally divided into probability sampling and non-probability sampling (Cohen et al., 2007, Ch.4). The former leads to the generalisation of research findings from the sample to the wider population, whereas the latter does not. In other words, a probability sample represents the wider population, whereas a non-probability sample does not.

Under the two umbrellas of sampling, there are a number of specific sampling techniques. Probability sampling methods include simple random sampling, systematic sampling, stratified sampling, cluster sampling, stage sampling, and multi-phase sampling (Cohen et al., 2007, Ch.4). Non-probability sampling methods consist of convenience sampling, quota sampling, purposive sampling, dimensional sampling, snowball sampling, volunteer sampling, and theoretical sampling *(ibid.*).

3.3.3.1 Sampling method

Given the limitation of a PhD study in terms of resources, a combination of convenience and stratified sampling was applied. Convenient samples are usually obtained for easy access. Conventionally, this approach is not probability sampling. However, the combination of it with probability sampling increases the statistical strength of the combined approach. Stratified sampling is a way of probability sampling in which the population is divided into "homogenous groups, each group containing subjects with similar characteristics" (Cohen et al., 2007, p. 111). With 'a useful blend of randomization and categorization' (ibid., p. 112), this sampling method allows the researcher to collect data that can not only be analysed quantitatively, but also be interpreted qualitatively, which fits a mixed-methods study well. Focusing on a specific stratum also helps limit the number of researchers and amount of time and other resources that a study may need.

A Chinese studying in England, the PhD student is interested in both international comparisons and what works in the teaching of mathematics. Thus, England and China naturally come to the fore as potential contexts for the study. Despite the initial consideration of convenience, the student still wants to come up with a sample that has statistical traits and thus can better inform research, practice and policy.

In order to compare like with like, it is considered reasonable to compare cities that are at the medium-to-high level in socio-economic terms in both countries. To limit the data to a manageable size, it is proposed that, in each country, data will be collected in one city representing the cities from the medium-to-high economic stratum of its home country. A focus on the medium-to-high level cities is due to the fact that China is still a developing country where the gap between rural and urban regions in terms of educational resources can be considerably big with the latter generally better equipped than the former, though the country has policies to support less developed regions (Peng et al., 2014).

Coming from Nanjing and pursuing a PhD in Southampton, the researcher thus attempts to understand whether the two familiar and of course convenient cities match the criteria. Subsequently, a series of national and municipal data searches has been carried out, which confirms the comparability of these two nations and cities in socio-economic terms.

Key information regarding the comparability and representativeness is presented in Chapter 1, which also works as a bird's-eye view of the research contexts. The following sections present the criteria for sampling schools and participants, detailed sampling procedure and the final sampling results.

3.3.3.2 Sampling criteria

At the heart of the stratified sampling method lie the criteria that define where and how participants are recruited. The EMT project's focus is on the population of teachers whose pupils (aged 9-10) are at the average level in medium-to-high level cities in each country. In this study, by *average level*, it means that two criteria are considered: (1) pupils' academic performance and (2) their socio-economic status (SES).

3.3.3.3 Sampling schools and participants

In Southampton England, sampling went through two steps: (1) locating representative schools and sending out invitations and (2) receiving final confirmations from participating schools and consents from teachers and parents/guardians.

Step 1: A list of potential schools was made through three sub-steps in September 2012:

(1a) Match the first sampling criterion (i.e. average mathematics performance) with the data on schools' mathematics performance in the Key Stage 2 Standard Assessment Tests (KS2 SATs) on the website of the *Department of Education* of the UK. Potential schools were located by comparing the average percentage of children achieving Level 4 or higher in mathematics in KS2 SATs over the past three years (2009 to 2011) in each school to the municipal average (77.3%).

(1b) Match the second sampling criterion (i.e. average SES) with the data on schools' intakes of free school meals (FSM) from the City Council's website. Schools were clearly clarified by the Council in terms of FSM take-up as Low, Medium and High, along which exact percentages of children taking FSM were also provided.

(1c) Cross-check the information collected in (1a) and (1b) resulted in locating seven schools with an average of 77% (SD = 2.1%) of children reaching mathematics Level 4 or higher and an average of 26.4% FSM take-up. Heads of schools were then contacted via email attaching with detailed information regarding the project. Several school visits were subsequently made on request in explaining details and exploring possibilities of participation.

Step 2: Through direct researcher-school communication via email and/or over meetings, headteachers from three schools, School EN-A, School EN-B and School EN-C, agreed to ask their Year-5 teachers and pupils to take part. Across three schools, children achieving Level 4 or higher mathematics averaged 76% (SD = 1.7%), with an average of 27.23% (SD = 2.02%) FSM take-up. Information sheets and consent forms were given to teachers and parents, and all Year-5 teachers and the majority of children's parents/guardians gave their consents/permissions, with one child opted out of tests and three opted out of classroom observations.

Judging by the sampling procedures and criteria, the final cohort of Year-5 teachers and pupils from the three schools chosen were representative of the average level of Year 5 teachers and pupils in Southampton.

In China, data on school performance and pupil SES were not openly accessible. Government claims to ensure equity during the stage of compulsory education (primary and lower secondary), which results in the prohibition of all forms of ranking or publicity of primary school performance. The prohibition is particularly strict in cities in China and is intended to prevent parents from rushing to buy properties in the catchment areas of certain 'best' schools.

Thus, in Nanjing, alternative measures were taken to locate the schools that met the criteria, which also involved two steps. Firstly, two professors, as experienced educational researchers from a well-known university in Nanjing, were asked to recommend schools that satisfied the two sampling criteria and that were willing to participate. These two professors were renowned to local schools and had been carrying out action research across schools in Nanjing and its surrounding areas. Later on, these two professors reassured that they had consulted an experienced *teaching research officer* (in Chinese: 教研员) from Nanjing Bureau of Education in the process of locating schools. As introduced in Chapter 1, teaching research officers in China work for the LEAs, and their duties are mainly inspecting teacher practice and performance. Overall, the two professors' and the teaching research officer's knowledge of school performance meant that their recommendations were close to the target cohort – the average level of Nanjing. Secondly, the headteachers from these schools were asked to invite Grade-4 maths teachers and their pupils to take part in the study on the basis of full consents.

These steps resulted in the participation of all maths teachers in Grade 4 from each participating school, so there was no selection of teachers within schools in either city. In Southampton, every Year-5 teacher was in charge of one class; in Nanjing, every Grade-4 teacher taught two parallel maths classes, and, depending on the date of observation, the class where a maths lesson was filmed got involved in all other data collection events.

3.3.3.4 Sampling results

After sufficient communication, research information was disseminated and well received, and participating consents were obtained. Initially, 10 teachers from three English schools and 12 teachers from five Chinese schools agreed to participate in the study along with their pupils aged 9-10. During the actual data collection process, all of the English teachers and their pupils (n = 236) participated in every data collection event; the initially 12 Chinese teachers and their pupils took part in all other events before the two tests, but three teachers and their classes from one of the five schools gave it up later on due to difficulties in timing the tests; for the remaining 9 classes (number of pupils in all events = 343) from four schools in Nanjing, two of them could not make the first test because of timing difficulties, so seven had sat the first test, and all nine classes (number of pupils in test two = 326) had taken the second test. Table 3.1 shows the final sample and the responding rates of each data-collection event. Details about participating schools and classes are presented in sections 1.4.3 and 1.4.4

Table 3.1 The final EMT sample and the response rates

	M1 & M4	*M2*	*M3*	*M5*	*M6*
EN pupils (n=236)	236 (100%)	231 (97.9%)	T1: 231 (97.9%) T2: 236 (100%)	–	–
CN pupils (n=343)	343 (100%)	343 (100%)	T1: 250 (72.9%) T2: 326 (95%)	–	–
EN teachers (n=10)	10 (100%)	10 (100%)	–	10 (100%)	10 (100%)
CN teachers (n=9)	9 (100%)	9 (100%)	–	9 (100%)	9 (100%)

The following section gives detailed descriptions of research methods that were applied.

3.4 Data collection & analysis methods

To answer the research questions, five data collection methods have been applied. The complete timetable for data collection can be found in Appendix F. These methods include classroom observations, interviews, focus groups, questionnaires, and standardised tests. In the process of data analysis, initially, results and findings are reported independently in partially answering research questions one and two, and later on, they are interconnected to address the research question three.

In order to clarify the description of data collection and data analysis methods and procedures, the aforementioned five research methods were further divided into six, with observations subdivided into structured and unstructured observations at the data-analysis stage. This section is therefore divided into six domains: *Structured classroom observations* (M1), *Questionnaires with teachers and pupils* (M2), *Standardised mathematics tests* (M3), *Unstructured classroom observations* (M4), *Interviews with teachers* (M5), and *Focus groups with teachers* (M6). These six domains are grouped under two umbrella titles, Methods – Part 1 and Methods – Part 2. The former contains M1, M2 and M3 which are quantitative methods, and the latter consists of M4, M5 and M6 which fall in the qualitative category. Detailed interconnections between methods may be found in the subsequent explanations of these methods.

These methods interconnected closely with one another. M1 and M3 provided data for the correlational analysis between teacher behaviours and pupil academic learning outcomes. M2 offered an overview of teacher background and beliefs and pupil background information. Overall, M1 to M3 addressed research question 1, M4 to M6 together addressed research question 2, and a combination of all methods addressed research question 3.

Methods – Part 1

The first part of the research methods includes structured classroom observations, questionnaires to teacher and pupils and standardised mathematics tests. For each method, the methodological literature will be reviewed, the utilisation of the method in the EMT project explained, and the corresponding instrument(s) described and justified, if any. Given the fact that lessons are at the heart of the project, the first method to be introduced is structured classroom observations. The review of methodological literature looks at observations in general, types of observations and structured observations. The final sub-section, *structured observations*, sees a synthesis of the literature and the application of this approach in the study.

3.4.1 Structured lesson observations (M1)

Observations in general

Observation is a widely adopted approach to collecting data for educational research (Punch, 2009, p. 153). It is regarded as a powerful research tool, since it enables the researcher to collect "live data from naturally occurring social situations" such as classrooms where teaching and learning take place (Cohen et al., 2007, p. 396). Unlike interview, the focus of observation is not on "what people say they do *but* what they actually do" (Gillham, 2008, p. 1). It can fit in either qualitative or quantitative paradigms or both according to the purpose of inquiry and the nature of data collected in each specific study (Wragg, 1999).

Observation also has its disadvantages. During the data collection process, observers are likely to affect the subjects and their behaviours, which is called the "observer effects" (Hiebert et al., 2003, p. 7). Observer effects might risk the quality of data collected and the reliability of findings (Cohen et al., 2007). Sufficient preparation can considerably reduce the observer effects, such as, arriving at the site earlier, communicating with school officials, the teacher and pupils, being transparent about the research itself, and gaining trust from the participants through friendly communication (LessonLab, 2012). At the data analysis stage, researchers might cherry-pick and prejudice what is observed, so they need to conduct the observation "in as objective a way as possible" to minimise biases (Bell, 2010, p. 195).

Types of observations

Observations are classified along five dimensions by Flick (1998, p. 137): (1) structured (QUAN) vs. unstructured (QUAL) observations; (2) participant vs. non-participant observations; (3) overt or covert observations; (4) observations in natural settings vs. artificial settings; (5) self-observations vs. observations of others. Cooper and Schindler (2001, p. 375) further add direct vs. indirect observations. In this study, observations are all *non-participant* and *overt* observations of *others* in *natural settings*. Lessons were observed and analysed in both *structured* (QUAN) and *unstructured* (QUAL) ways. The former offered a systematic evaluation of teacher behaviours, classroom activities, and pupils' time on task; the latter provided the researcher freedom to explore the lessons thoroughly and open-mindedly. The focus of this section is on structured observation and its application in the study.

Structured observations

A structured observation is a way of generating numerical data from either artificial or natural settings through a survey of pre-decided factors either by frequency or on rating scales in a given period (Cohen et al., 2011; Wilkinson, 2000; Wragg, 1999). Such numerical data offer opportunities to compare results across settings (Cohen et al., 2007). The focus of structured observation is pre-determined rather than naturally emerging during or after the observation (Bell, 2010).

Traditionally in a structured observation, observers usually sit at the corner of the observed setting, observe and record frequencies of or rate levels of focused events or factors on an observational schedule. The observers have to react and scan through many categories 'at speed', which demands longer pre-observation training for the observer(s) to be "proficient and consistent in entering data" (Cohen et al., 2011, p. 459). If there is more than one observer, then training sessions can also help achieve inter-rater reliability between observers to make sure data are entered consistently in the same categories (ibid.).

The use of video recording in observations

Along with the development of image technology, video-recorded observations are increasingly applied in a wide range of educational research, especially in research on teaching practices. Ulewicz and Beatty (2001, p. 8) noted that video recording was able to capture more details in a shorter time than any other traditional method for collecting data in classrooms. In the report of TIMSS 1999 Video Study, Hiebert et al. (2003) also pointed out the advantages of using video in studying classroom practices:

- Video enables the study of complex processes;
- Video increases inter-rater reliability, decreases training difficulties;
- Video enables coding from multiple perspectives;
- Video stores data in a form that allows new analyses at a later time;
- Video facilitates integration of qualitative and quantitative information;
- Video facilitates communication of the results.

(Hiebert et al., 2003, pp. 5-6)

For this study, three video cameras were utilised to record lessons (n=19) in both countries. Participants were all informed that lessons were to be video-recorded. The researcher generally arrived at schools half an hour earlier to get familiar with the teacher, pupils and the classroom. Cameras were placed at suitable places where ideal pictures could be captured and where pupils were less likely to walk around. Details about placing cameras are explained as follows.

Field placement of the cameras

In order to shoot the lesson with the teacher's, pupils' and researcher's perspectives all considered, three cameras were set up in every classroom. These included two smaller handycams (Cam #1 and Cam #2) and a larger-size professional camcorder (Cam #3). Each camcorder was attached to a tripod.

Cam #1 and Cam #2 were set up pointing to two randomly chosen groups of pupils to capture their activities throughout a lesson. In most cases, they were placed at each side of the front board where the teacher generally gave demonstrations to the whole class. Once the two small handycams were set up and switched on before the lesson started, there would be no other operations on them throughout the lesson. The researcher was then free to go and concentrate on the operation of Cam #3. Cam #3 was placed near the wall *opposite* to the front board to offer wide-angle view of the whole class. Through Cam #3, the researcher was trying to view the lesson from both the points of view of pupils and an outsider – the researcher. Mostly, it kept following the teacher and occasionally zooming in to see details, for example, about the teacher's marking on a pupil's work.

Lesson videos were analysed in structured and unstructured ways. As follows, instruments applied for structured observations are to be introduced.

Instruments for structured observations

Lesson videos were measured against two existing observational systems developed by two international research teams who have conducted educational effectiveness research cross-nationally.

System A, Opportunity to Learn, (see Appendix B) was adopted from the International School Effectiveness Research Project (ISERP) (Reynolds et al., 2002a). ISERP was a SER study across nine countries and four continents (America, Asia, Australia and Europe). It has effective school factors and effective teacher factors in different contexts as the first of its three research questions. The ISERP Research Team designed the observational system, Opportunity to Learn (OTL), as one of the nine optional instruments and protocols for research at pupil and teacher levels.

OTL has a particular focus on percentages of pupils' time on task (in the right column) and percentages of time allocation to five different activities (in the left column). Plus pupil time on task, OTL has a total of six measures.

The first thing for evaluating a lesson under OTL was to divide the lesson into small segments and code these segments into the five OTL activity categories:

- *Whole class interaction*
- *Whole class lecture*
- *Individual/group work*
- *Classroom management*
- *Partial class interaction* (replacing testing/assessment)[1]

[1] Partial class interaction was repeatedly discovered in almost all English classrooms during the initial lesson data analysis. Testing/assessment behaviours were always found netted in either whole class or partial class interactions in English lessons and with whole class interactions in Chinese lessons. It was thus decided that "partial class interaction" would replace "testing/assessment" as the fifth component of OTL to better describe the quantity of teaching.

While observing a lesson video, firstly, the researcher jotted down the starting and ending time points of each activity and coded it into one of the five activities. Secondly, the time span of each coded activity was calculated by subtracting the starting point from the ending point. Thirdly, all time spans for each specific OTL activity were summed up to get the total time on that specific activity in the lesson. Finally, it is by dividing the lesson duration by the total time on each specific activity that the percentage of time on that activity was calculated.

The second thing using OTL is to count the number of pupils on task every five minutes, divide the sum of pupils by the on-task number and calculate a mean to represent each lesson's time on task.

System B was the protocol for the International System for Teacher Observation and Feedback (ISTOF, see Appendix B), designed, applied and validated in two empirical studies across over 20 countries by Teddlie et al. (2006). The development of the ISTOF protocol was partially based on methodological lessons learned from ISERP and its observational systems – the Virgilio Teacher Behaviour Inventory (VTBI) and the Quality, Appropriateness, Incentive and Time (QAIT) Rating Scales.

The main purpose of ISTOF was to "develop an internationally valid system for assessing teacher effectiveness" (ibid, p. 565). The reasons for developing ISTOF include:

- the necessity to develop an observational system that could travel successfully across countries,
- the need of such an instrument for the development of TER literature among countries,
- the lack of a system that could provide post-observation feedback to teachers among countries,
- the necessity of a more advanced instrument for future international EER studies

Twenty-five country teams contributed to the development of ISTOF. The components and indicators of effective teaching were generated among these countries initially in 2004-2006. From 2006 to 2007, the protocol was field-tested, psychometric indices were generated, and guidelines for observational feedback were developed.

The original ISTOF protocol contains seven domains: *Assessment and Evaluation, Differentiation and Inclusion*, *Clarity of Instruction, Instructional Skills*, *Promoting Active Learning and Developing Metacognitive Skills*, *Classroom Climate*, and *Classroom Management*. Within each domain, several subdomains (indicators) were listed, which in turn were subdivided into various teacher behaviours. Overall, there are 45 teacher behaviours measured on a 5-point Likert scale, ranging from “1 = strongly disagree” to “5 = strongly agree”.

In this study, one domain named *Differentiation and inclusion* (items #5 to #8) and one differentiation-relevant item (#20) from the *Instructional skills* domain were taken out. The decision was made *not* simply because differentiation was unimportant but because all Chinese schools were mixed settings within and beyond classrooms. Chinese teachers might be dealing with the ability differences among pupils in a different way from what the five items have described, hence the decision. Future research might be focusing on this domain to develop new items. In addition, the picture of differentiation seems to be more confusing especially when this study shows a rather smaller performance gap in China than that in England (see the *Standardised mathematics tests* section in this chapter). It thus poses questions as to why teaching in mixed-ability settings generated small performance variance in Chinese schools and why English teachers devoted huge effort into differentiating teaching but created a much wider range of learning outcomes.

Further analysis

Two observational systems supplemented with each other, in that they provide both macro and micro evaluations on a lesson. As lessons were all video-recorded, the traditional *on-site* ticking and scaling on a structured schedule were postponed for off-site analyses on lesson videos. Before formal analyses, the research student and the supervisor carried out trial observations on four lessons, two from China and another two from England. Cohen's Kappa was employed to calculate inter-rater reliability. Cohen's Kappa of 0.78, 0.81, 0.84 and 0.88 were attained in the four sets of ratings respectively. After completing each set of ratings, the two raters discussed the items where disagreements lied so as to facilitate the student's more accurate understanding and using of the instrument. An average score of Cohen's Kappa at 0.83 (> 0.7) suggested that the inter-rater reliability had been achieved. The student was then ready to formally evaluate lesson data independently.

The formal process involved ratings and data input. Raw results from both systems were input into Excel and further analysed with the aid of SPSS. Teacher variables consisted of percentages of *pupil time-on-task* and *teacher time-allocations* to five types of classroom activities measured with the ISERP system, and the 40 *teacher behaviours* measured with the ISTOF system. Details are reported in Chapter 4.

3.4.2 **Questionnaires to teachers and pupils (M2)**

Subjects' traits in educational, psychological and social research are generally *latent traits* (Punch, 2009, p. 239) or *latent variables* (Muijs, 2011, p. 57) that are not directly observable. The latent traits or variables can, however, be inferred by the results of the measure of observable indicators that are believed to be closely connected with those traits (Punch, 2009). A questionnaire is one of many ways to measure these indicators.

As a widely applied survey tool, the questionnaire's beauty lies in its ability to collect information in an often highly-structured way and therefore generate numerical data (Cohen et al., 2011). This does not mean that the questionnaire consumes less time. On the contrary, constructing it costs more time than the subsequent dissemination and analysis. The instrument has to be established, tested, modified and retested as necessary until both reliability and validity have been attained at an optimal level (Punch, 2009).

Because of the time and resources required to construct a questionnaire, it is worth careful consideration whether to construct a new instrument or use an existing one. If there is an instrument that has already been constructed with regard to the same trait(s)/variable(s), it is argued that using the existing one is a wiser choice in at least two ways: (1) previous studies have helped in increasing our knowledge on the properties of the instrument; (2) research results can be easily "compared, integrated, and synthesised among studies where this instrument has been applied" (Punch, 2009, p. 243).

For this study, two existing questionnaires on similar areas had been discovered during the process of literature review. One was the OECD's survey (2009a), *TALIS 2008*, which shared the central variables that this study intended to survey through a teacher questionnaire: teacher beliefs, teacher self-efficacy, teacher self-evaluation on teaching practice, and the school climate. Moreover, the survey had been done internationally, which fit well with the cross-national feature of the current study. The other was the *Pupil Perception Survey for Elementary Pupils*, from the Measures of Effective Teaching project (Bill & Melinda Gates Foundation, 2012c) where TER was central to its focus, which made the questionnaire fit perfectly with the attempts of the EMT study to collect pupils' perceptions on TER. Thus, there seemed no need for the researcher to reinvent the wheel for this study. The two instruments were adopted and applied in this study after minor amendments.

The application of the method in this study

Teacher Questionnaire

There were 43 questions containing 174 items in the original TALIS 2008 Teacher Questionnaire (OECD, 2009b) where questions were grouped under five titles (in brackets: the numbers of items): *Background Information* (n=13), *Professional Development* (n=27), *Teacher Appraisal and Feedback* (n=52), *Teaching Practices, Beliefs, and Attitudes* (n=56), and *Your Teaching at Particular Class at This School* (n=36).

The focus of the TALIS 2008 Teacher Questionnaire was on teacher effectiveness and efficacies, so the latent traits that the original survey intended to investigate were the same as the EMT project proposed to do. Nonetheless, before the formal application of the questionnaire in the study, it is necessary for the researcher to make a few adjustments to the original to meet specific requirements of the current study. Details regarding the revising process are as follows.

The revision of the Teacher Questionnaire

There are three aspects that were considered for either small amendments or deletion. First, any questions regarding teaching were clearly re-defined as *mathematics* teaching when necessary, and the default schooling phase was changed from secondary to primary to suit the study's target level.

Second, the names of education degrees were all changed from the UNESCO's ISCED levels to everyday terminologies that teachers were familiar with. For example, question #7 asked teacher respondents the highest level of education they completed, providing choices such as "Below ISCED Level 5", "ISCED Level 5B", and so on (OECD, 2009b, p. 4). As a result of revision, the original choices were replaced by choices such as "Bachelor degree", "Masters degree", and "Doctoral degree" (see Appendix E: Teacher Questionnaire, p. 2). The subsequent fieldwork proved that this amendment made sense to all teacher participants.

Third, as the TALIS 2008 was surveying teachers from all disciplines and teaching at secondary levels, the original questions #34 – #37 asked teachers what specific subjects they were teaching and at what levels. However, these questions were meaningless for the EMT project, as all teacher participants were primary teachers. Thus, the four questions were abandoned. In addition, the original items #33d and #33j were about social studies and vocational studies provided in secondary schools, instead of primary schools, so they were deleted as well. In total, six items were removed from the original instrument.

Table 3.2 Overview of the revision of Teacher Questionnaire

Item/Variable types	*Original Version*	*Revised Version*
Likert scale variables	144	144
Other ordinal variables	7	6
Dichotomous(Yes/No) variables	19	16
Continuous variables	11	11
Nominal variables	3	1
Total	*184*	*178*

A comparison of types and numbers of items between the original and revised versions is shown in Table 3.2. Overall, the revised questionnaire consisted of 178 variables, of which the majority (n=144) were Likert scale items. The English version was translated into Chinese, and then the translated version was translated back into English again to compare with the original. The comparison suggested that the Chinese translation contained the information the original instrument intended to tell. Then, a Chinese teacher (non-participant) was asked to review the translated questionnaire to check whether it made sense to Chinese teachers. After this step, the Chinese version was finalised.

Pupil Questionnaire

The questionnaire for pupils was adopted from the MET project (Bill & Melinda Gates Foundation, 2012c). To measure pupils' perceptions in seven conceptual domains, two questionnaires were developed by Harvard researcher R. Ferguson, one for primary and the other for secondary school pupils (Bill & Melinda Gates Foundation, 2012a, 2012d). The questionnaire was intended to measure seven C's – *Care, Control, Clarity, Challenge, Captivate, Confer* and *Consolidate* – seven indicators of what the MET team believed as effective teaching. Sharing the same purpose of surveying pupils' perceptions of teaching effectiveness, the MET questionnaire for primary pupils was adopted for this study after minor amendments.

The revision of the Pupil Questionnaire

The pupil questionnaire involved three aspects of revision. First, comprehensive amendments were done throughout the original instrument mainly to change American English into British English, such as changing *students* into *pupils*, and to change the discussed subject content from ambiguity to mathematics. Because the MET project was surveying teaching and learning in two subjects, mathematics and reading, the questionnaire asked pupils questions in quite a general way. For example, in the original questionnaire, item #1 asked, "I like the ways we learn in this class", whereas it was revised as "I like the ways we learn in maths lessons in this class" to keep the survey firmly focused on the target subject area of this study.

Second, a number of items have been abandoned for certain reasons. There were seven original items, #45, #47, #48, #55, #58, #72 and #81, which have been abandoned for both English and Chinese versions of the pupil questionnaire. Items #45, #47, and #55 were about reading and writing, which were not relevant to mathematics, and there seemed to be no possibilities to further develop them into mathematics-relevant questions. Items #48 and #58 were repetitive items which had been asked elsewhere. Item #72 asked about the age group that pupils belonged to, which was unusable for the EMT study as pupil participants were clearly all aged 9-10. Original item #81 questioned what adults were with the pupils when they did the questionnaire, which did not need to be answered because the researcher delivered the questionnaire papers from class to class in this study. In addition, for the Chinese version, item # 74 was also abandoned. It asked whether the pupil's family spoke English at home, but Chinese do not speak English at home, which made this item somewhat nonsensical in the Chinese context.

Third, an original item #78 (#72 in the finalised English and Chinese versions) has been revised in two ways to suit the corresponding context. This multiple-choice question was intended to ask pupils' ethnicities. Apparently, each country has its unique structure, types, and names of ethnic groups. The original choices provided were names of typical American ethnic groups, which do not apply to either country involved in this study, so the original choices were deleted and replaced with a set of English choices and a set of Chinese choices for each version to fit well into each country's context. Nevertheless, during data analysis, it was the proportion of the ethnic group in the respective country that defined whether the pupil, who chose a specific answer, was from an ethnic majority or minority.

Table 3.3 Overview of the revision of Pupil Questionnaire

Item/Variable types	*Original Version*	*Revised Version*
Likert scale items	67 (5-point scales)	62
Multiple-choice items	14	12 (11 for China)
Total	*81*	*74 (73 for China)*

In Table 3.3, the finalised Pupil Questionnaire contained 74 questions (variables) for the English version and 73 for the Chinese version, with the majority (about 6/7) as Likert scale items.

As aforementioned, both surveying instruments were translated into Chinese, and the Chinese versions were translated back to English to double-check whether the Chinese instruments carried the same information. After the translation and back-translation process, a Chinese primary pupil, who is the relative of the researcher, was asked to proofread the translated questionnaire. Afterwards, a few words had been revised to ensure the language made sense to children in Chinese primary schools without changing the fundamental meaning of the language. Thus, a final Chinese version of the pupil questionnaire was produced.

Unexpected issue emerging during the pupil questionnaire analysis

Data collection went as smoothly as expected, with a total of 576 children from 19 classes involved and a response rate of 99.7% (574 completed). However, during the data analysis process, only 34 domain-specific items (i.e. labelled with the 7 C's) were found in Table 1 in a MET's initial report (Bill & Melinda Gates Foundation, 2012b, p. 12), and two of them were about reading and writing and were already excluded from the EMT version. No answer had been received after several attempts to approach the author, and there were no further publications available regarding the pupil survey from their project. Thus, only the background information and 34 classified items were analysed and presented in the thesis to give an overview of the pupil cohort and their perceptions of maths teaching and learning in the 7 C's.

Data analysis methods of questionnaires

Descriptive statistical analyses with the aid of SPSS

Both questionnaires were analysed in a similar way. First, questionnaire papers were transformed one by one into digital forms on SPSS spreadsheets with questionnaire items as columns and individual respondents' answers to all items as rows. Second, questionnaire answers were coded as numbers in the value labels cells in the variable view of SPSS spreadsheets. Third, items in both questionnaires were coded into different domains. The domains for the teacher questionnaire items were clearly presented in the instrument. Pupil background information was grouped under three titles: pupil gender and ethnicity, pupil family background and pupil-perceived schooling.

The majority of items in both questionnaires were ordinal variables, and a small number of them were nominal variables. As the purpose of analysing the questionnaires were to compare responding frequencies between countries (a nominal variable), the statistical comparisons were either between "*a nominal variable and an ordinal variable*" or between "*two nominal variables*" (Muijs, 2011, p. 99). Thus in SPSS, the cross-tabulation was applied to illustrate cross-national comparisons on questionnaire variables, whereas the chi-square test was conducted to see whether the differences between countries were obtained by chance. In addition, the measure of phi was conducted to reveal how strong the relationship was between the independent variable (in this case, nationalities) and the dependent variables (in this case, questionnaire responses) and if there was a statistically significant relationship between the two.

3.4.3 Standardised mathematics tests (M3)

As explained in the previous section, like the questionnaire, the standardised test is also a way of measuring latent traits or variables indirectly. Whether a pupil has learnt mathematics well or not can not be measured by "plug[ging] directly into" his head to see how good his mathematics ability is (Muijs, 2011, p. 57). It can be inferred by observable indicators. A set of items can be reasonably sampled from "among [the] theoretically infinite set of observable indicators" (Punch, 2009, p. 239) to form an instrument for the measurement of the pupil's mathematics level.

The role of the tests in the study

Central to the TER lie the correlations between teaching behaviours and pupils' academic performance. This has rarely been conducted comparatively in the proposed contexts, England and China. For this study, two standardised mathematics tests were proposed to examine pupils' mathematics learning outcomes at two points in time with a same set of test items. The reason for testing twice was an attempt to test learning gains over time, in addition to the initial research target – evaluating mathematics learning outcomes. However, the test instrument, as will be introduced subsequently, was adopted from TIMSS which aimed at testing learning outcomes at points in time. Thus, the test paper might not be able to reveal reliable learning gains over time, hence an attempt. Test results on two occasions were correlated with teachers' effectiveness scores measured against two observational systems.

The reasons for using TIMSS 2003 items

Again, as was the case with questionnaires, it would be relatively time and energy consuming for a PhD project to construct a new test instrument, with reliability and validity established across countries, when there were several other methods also awaiting implementations. Adopting an existing high-quality instrument was an ideal choice. In this project, it had to be a parametric, norm-referenced, domain-referenced, and researcher-produced test (Cohen et al., 2007, Ch.19). After careful consideration, one of TIMSS past papers for Grade 4 (aged 9-10) was seen as a better choice than British SATs papers and other commercial test papers. As follows is the rationale for choosing TIMSS 2003 as well as the researcher's train of thought in editing a set of paper out of the TIMSS 2003 released item pool.

Firstly, international reputation is considered. TIMSS and PISA are two long-standing brands of international assessments testing subjects including mathematics. PISA is testing at late secondary phase which does not match to the target population of pupils in this study – pupils aged 9-10 at the late primary phase, whereas TIMSS is testing at both primary and secondary levels. Hence, test papers from the latter are more likely to fulfil the research purpose of the study. For the primary assessment, the target cohort in TIMSS is Grade 4 – the same as this study's target age group, Grade 4 in China and Year 5 in England.

Secondly, TIMSS assessments were established collaboratively by mathematics and science educators and development specialists (e.g. see the case of TIMSS 2003 in Martin et al., 2004, p. 8), and the reliability and validity of the assessments were attained internationally.

Thirdly, TIMSS past items are partially available from 1995 to the now downloadable 2011 items, but which year to choose remained a question. After TIMSS 1995 and its repeat study TIMSS 1999, it is considered that TIMSS 2003 would be a paper best represent international trends through three cycles of assessments as TIMSS team claimed to do.

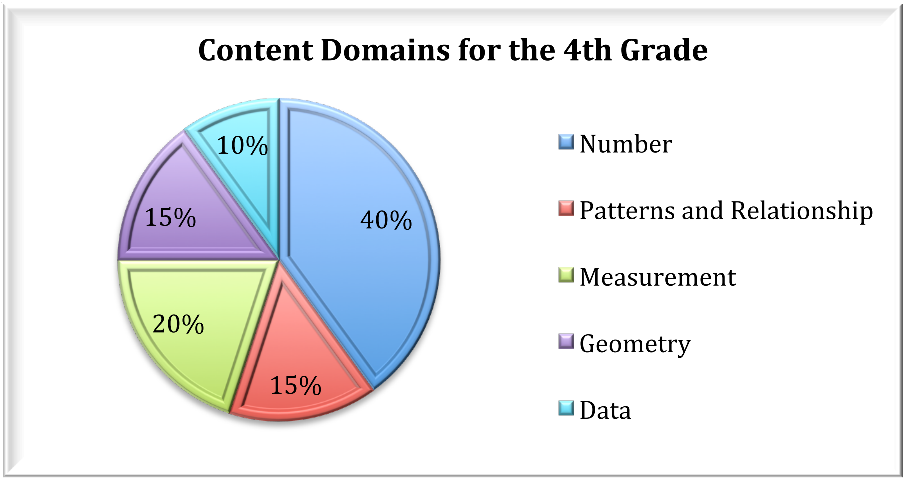
Last but not least, parallel to TIMSS 2003 released items, IEA also published a set of marking guidance which clearly labelled each item with the cognitive domain, content domain, and topic it belonged to. In addition to providing the correct answer, an international benchmark was also presented for each item (IEA, 2007).

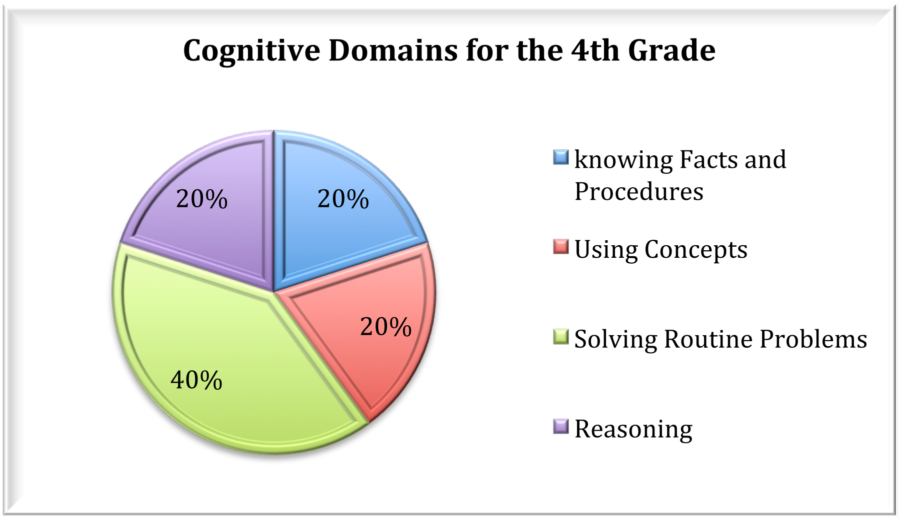
All reasons seemed to point the researcher in the right direction. There were, however, two drawbacks: (1) IEA had only released a half of each cycle's item pool, keeping the other half for measuring trends in future; (2) a cycle of TIMSS tests consisted of 12 booklets where both science and mathematics items were distributed proportionally with regard to its framework-defined percentages of content and cognitive domains and the expected ratio of item types. Therefore, with TIMSS 2003 released items in hand, there was a crucial task for the researcher to do – to reconstruct a booklet that contained items in ways that could match the newly constructed instrument well with the TIMSS 2003 assessment framework.

The formulation of the test paper

For the EMT project, strict measures were taken in choosing items and embedding the psychometric properties of TIMSS 2003 in the newly organised test paper.

The following aspects were the centre of attention: (1) percentages of various content domains and cognitive domains defined in the TIMSS 2003 Assessment Framework, (2) a balanced coverage of main topics within each content domain, (3) percentages of various item formats, (4) timing the test and (5) social and cultural adaptability of the test.





Source: Martin et al. (2004, p.29)

Figure 3.4 TIMSS 2003 Assessment Framework (G4 Maths)

Content domains and cognitive domains

TIMSS 2003 Assessment Framework clearly defined the percentages of content and cognitive domains. Figure 3.4 shows the distribution of all domains in the TIMSS 2003 Assessment Framework for the 4th-Grade mathematics. As will be explained in the section *Timing of the test*, the average time for each item is about one minute. Choosing 40 items to form a test paper means that the test would take a period of 40 minutes. The amount of time for a period in primary schools in England is about 50 minutes, and that in China is about 40 minutes. A 40-minute test seems reasonable for both sites. Bringing all of these circumstances together, it was decided that 40 items were to be chosen from the TIMSS 2003 released item pool.

Table 3.4 Item numbers by content domain in the EMT test

Content Domains	*Released No.*	*Expected No.*
Number	37	16
Patterns and Relationships	7	6
Measurement	14	8
Geometry	13	6
Data	8	4
Total	79	40

Table 3.4 shows a comparison of the numbers of items in various content domains between TIMSS released item pool (Released No.) and the then proposed EMT test paper (Expected No.).

Table 3.5 Distribution of items in content and cognitive domains

	Knowing facts and procedures	*Using concepts*	*Solving routine problems*	*Reasoning*	Overall No.	*Expected No. (percentages)*
Number	4	4	6	2	16	16 (40%)
Patterns and relationships	1	1	3	1	6	6 (15%)
Measurement	1	1	3	3	8	8 (20%)
Geometry	2	0	3	1	6	6 (15%)
Data	0	1	2	1	4	4 (10%)
Overall No.	8	7	17	8	40	40 (100%)
Expected No. (percentages)	8 (20%)	8 (20%)	16 (40%)	8 (20%)		40 (100%)

It is evident in Table 3.5 that a 40-item test could simultaneously satisfy both the ratio of items from different content domains and the ratio of items from various cognitive domains, if operated carefully. The light shaded column and row show the numbers of included items in the EMT test by domain. These numbers are either equal to or very close to what were expected and defined by the TIMSS 2003 Theoretical Framework.

Table 3.6 Distribution of main topics in tested content domains

Content domains	*Main topics*	*Included No.*	*Released No.*
Number	Whole numbers	12	25
	Fractions and decimals	3	12
	Ratio, proportion, and per cent	1	1
Patterns and Relationships	Patterns	3	3
	Equations and formulas	2	3
	Relationship	1	1
Measurement	Attributes and units	2	5
	Tools, techniques, and formulas	6	9
Geometry	Lines and angles	1	1
	2D and 3D shapes	3	8
	Congruence and similarity	1	2
	Location and spatial relationships	1	1
	Symmetry and transformations	0	0
Data	Data collection and organisation	0	0
	Data representation	3	5
	Data interpretation	1	3
Total		40	79

Main topics of content domains

Apart from the previously discussed strategies on building up the test paper, during the item screening process, attention was also drawn to maintaining an optimal balance between the main topics in each content domain.

Table 3.6 shows a comparison between the numbers of included items and those of released items that belong to a variety of main topics respectively. Apparently, within all domains and main topics, the numbers of items that were intended to be included in the EMT test are all smaller than those of released items. This further proves that, with the number of items from each domain fixed, it is possible to choose 40 items that are nearly evenly distributed to all of the main topics of each domain.

Item formats

TIMSS 2003 for the Fourth Graders (G4) consisted of 161 mathematics items and 152 science items (Martin et al., 2004, p. 10). The "*constructed-response items* : *multiple-choice items"* ratio was between 1:2 and 2:3 (ibid.). Table 3.7 shows that the distribution of the two item formats in the EMT test paper was nearly equal to that in TIMSS 2003.

The timing of the test

The original 161 Grade-4 mathematics items in TIMSS 2003 were divided among 14 blocks, and so were science items, with each block containing around 12 items

(Martin et al., 2004, p. 53). Each booklet included six blocks which comprised approximately 72 items of mathematics and science. The time for completing a test booklet was 72 minutes, so, on average, each item was given one minute. The EMT test paper contained 40 items, so the test time was 40 minutes.

Table 3.7 Distribution of two formats of items in the EMT test

Item formats	*Achieved Percentages in EMT (Item number)*	*Defined percentages in TIMSS 2003 **
Multiple choice items	55% (22)	54%
Constructed response items	45% (18)	46%
Total	*100% (40)*	*100%*

* Source: Martin et al. (2004, p. 45)

The social and cultural adaptability of the test

The released items were in American English. After the test items had been screened and chosen, the main effort was given to transforming American English into British English for the English version of the test. Then, the English version of the test paper was translated into Chinese and translated back to English again to ensure the accuracy of the Chinese version. In word problems where names were involved, English names were all replaced with typical Chinese names to make the test more adaptable to the target context. Then two Chinese pupils were asked to proofread the translated test paper and see whether all items made sense to them. The result was optimistic – nothing inappropriate. Thus, the Chinese version of the test paper was ready to use.

Summary

With all the five aspects of measures undertaken, the standardised mathematics test paper was finally constructed, carrying the psychometric properties of TIMSS 2003 introduced in this section. It contained 40 items, so pupils aged 9 to10 were expected to complete the test in 40 minutes. The two mathematics tests were both pen-and-paper tests implemented during the second half of the school year 2012-2013. The amount of time between two tests was controlled for ten school weeks, and specific times and dates were suggested by teacher participants. As aforementioned in ***sampling processes and results*** in Section 3.3.3 because of timing difficulties, teachers of Classes CN5 (40 pupils) and CN6 (36 pupils) were not able to find a proper time slot for Test 1, which meant an increase of the number of Chinese pupils from Test 1 to Test 2 (Table 3.1). Because Chinese schools are all mixed-ability within and across classrooms and schools, the absence of two classes in Test 1 does not necessarily influence the overall results.

Data analysis methods of the test

Marking and analysing the test

The two paper-based tests were marked and double-checked by the researcher following the TIMSS 2003 marking guidance (IEA, 2007). The score of every paper was the percentage of correct answers the pupil achieved. Meanwhile, the percentages of correct answers that pupils achieved in each content domain and each cognitive domain were also calculated. The test results were recorded in SPSS spreadsheets where classes (teachers) and countries were set as independent variables. The classroom-level means of pupil performance was seen as a dependent variable. Independent-samples t-test was conducted to evaluate the significance of the difference between the means of pupils' test results in the two countries. To test how strong such a difference was, Cohen's *d* was calculated subsequently.

Correlating teacher behaviours with the test results

Mean scores of each of the two tests were correlated with the mean grades or percentages of teacher behaviours measured against the two observational systems. Pearson's *r* correlation coefficient was calculated with teacher behaviour scales/percentages as *x* and pupil test performance as *y* at the classroom level across countries.

Methods – Part 2

The second part of the research methods includes unstructured observations, interviews, and focus groups. The three methods were applied to collect multiple perspectives that all participants and the researcher hold with regard to maths lessons in two countries. The teacher who delivered the lesson gave personal explanation of her/his general beliefs and commented on her/his own lesson.

From all lessons observed in each country, one was chosen for group discussion in focus groups. As may be found in the section 3.5.6 *Focus groups with teachers*, all teacher participants were divided into four groups, two in each country, to collaboratively comment on the two lessons – one from China and the other from England. Thus, regarding two out of all lessons observed in this study, there were two extra strands of perspectives: native colleagues' views and foreign colleagues' views. Finally, the researcher offered her interpretation of maths teaching in two countries through unstructured observation.

The utilisation of these methods makes multiple voices heard and the study more democratic. Participants' opinions could deepen our understanding of insiders' views. As the *players*' willingness to make change is the sole lever to make a real change in classrooms, the first step is to hear and make sense of their thoughts. In addition, the two domains of findings from MM1 and MM2 were linked with each other after independent analyses under each paradigm, i.e. positivism and interpretivism. Such interconnection between MM1 and MM2 analyses plays an essential role in the final findings of the whole study, in that it helps bring together the researchers (not just the author), practitioners, policymakers, and other audiences from the wider world for possible collaboration for better mathematics teaching and learning in future.

In the following sections, the use of unstructured classroom observations, interviewing, and focus groups will be explained with regard to what the methodological literature has offered, how the three methods were applied in this study and how data were analysed accordingly.

3.4.4 Unstructured lesson observations (M4)

As introduced earlier, lesson data were analysed in both qualitative and quantitative ways, and structured and unstructured observations were conducted. This section is to focus on unstructured observation and its application in this study. There were overlaps between the two types of observations for this study, so those that have been introduced in Section 3.4.1 on *structured lesson observations* will not be repeated here.

Unstructured observation is generally adopted when the research purpose is clear but categories and concepts of data are not decided until they emerge during the analysis (Bell, 2010). In comparison with structured observation, it is more open-ended and natural (Punch, 2005) and more capable of revealing "larger patterns of behaviour, more holistically and more macroscopically" (Punch, 2009, p. 155). The subsequent findings might point the researcher towards a hypothesis (Bell, 2010). Unstructured observation is, however, not easy to manage (ibid.) and the analysis of data is more demanding (Punch, 2009).

Data analysis methods

Unstructured observation is conducted and analysed in a qualitative way (ibid.). Qualitative data analysis methods are diverse "because there are different questions to be addressed and different versions of social reality that can be elaborated" (Coffey & Atkinson, 1996, p. 14). It is hard to point out "a single right way" for analysing qualitative data (Punch, 2009, p. 171). Despite the diversity of analysis methods, qualitative data should be interpreted "in a rigorous and scholarly way – in order to capture the complexities of the social worlds we seek to explain" (Coffey & Atkinson, 1996, p. 3). The process of qualitative analysis generally involves coding and note taking which enable the researcher to generate abstract patterns from concrete observable phenomena (Punch, 2009, Ch. 9).

In terms of how to organise qualitative data, Cohen et al. (2007, pp. 467-468) suggested five ways: (1) by groups, (2) by individuals, (3) by a particular issue, (4) by research question and (5) by instrument. For unstructured analyses of lessons, mainly the first way was chosen so as to show characteristics of maths teaching in each country; typical snapshots from individuals' lessons were also drawn.

3.4.5 Interviews with teachers (M5)

The interview is a data-collection approach commonly used in educational research as "a conversation with a purpose" (Dexter, 1970, p. 123). It is the word *purpose* that distinguishes it from a common conversation in daily life. Interviewing and questionnaires are both approaches for surveying. Nonetheless, the former gets participants more involved, so it is more helpful in collecting rich and thick information than a questionnaire (Oppenheim, 1992). Interviewing is a way of "accessing people's perceptions, meanings, definitions of situations and constructions of reality" (Punch, 2009, p. 144). The application of interviewing prevents researchers from seeing participants as "simply manipulable and data as somehow external to individuals" (Cohen et al., 2007, p. 349). Interviewing enables researchers to "get the story from the point of view of the participant" (Lichtman, 2010, p. 139). As cited in Punch's (2009, p. 144) work, Jones puts the interview in a fairly nice way:

> In order to understand other persons' construction of reality, we would do well to ask them … and to ask them in such a way that they can tell us in their terms (rather than those imposed rigidly and a priori by ourselves) and in a depth which addresses the rich context that is the substance of their meanings.
>
> (Jones, 1985, p. 46)

As commonly seen in the typologies of many other research methods, different methodological researchers divide the types of interviews under various names in different ways (Cohen et al., 2007). Informational conversational interviews, interview guide approaches, standardized open-ended interviews, close quantitative interviews were classified by Patton (1980). Lincoln and Guba (1985) included structured interviews among other types. Semi-structured interviews and group interviews were added by Bogdan and Biklen (1992). Oppenheim (1992) added exploratory interviews. Later on, types of interviews were also defined as standardized interviews, in-depth interviews, ethnographic interviews, elite interviews, life history interviews, and focus groups by Lecompte and Preissle (1993). Noticeably, Minichiello et al. (1990) make things less complicated by organizing various types of interviews under the continuum model. This model clustered multiple types of interviews into three major categories: structured interviews, focused or semi-structured interviews, and unstructured interviews.

The application of interviewing in the study

In the EMT study, teachers did not only carry out lessons for external evaluations on the effectiveness of their teaching, but also were invited to talk about their own teaching beliefs, give their own comments on the observed lessons while watching the lesson video, and talk about their thoughts about mathematics teaching in the other country. Interviews with teachers did not just aim at collecting teachers' views, but rather provided rich information about why they were teaching in the way they were. The use of lesson videos helps clarify the focus of interviewing. The advantage of video-stimulated interview is evident in a number of previous studies, such as Learner's Perspective Study (Clarke et al., 2006) and Preschool in Three Cultures (Tobin et al., 2009; Tobin et al., 1989). Similarly, during the interviewing processes of this project, videoed lessons served as stimuli for teachers to travel through time to revisit their lessons and comment on any event at any convenient time point.

The type of interview applied

Given the focus of the study, interviews are intended to collect teachers' personal beliefs on teaching effectiveness and their comments on the quality of teaching. In order to guide the conversation between the teacher and the researcher towards the research focus, "a framework is established", containing questions relevant to the focus (Bell, 2010, p. 165). Though these questions were all pre-defined, the answers to them were completely open, hence making the interviews *semi-*

structured. In order to ask questions when each lesson was still fresh in a teacher's mind, every interview was scheduled for the same day or the day after the observation.

The approach to recording interviews

The use of video recording in interviews might help capturing non-verbal languages, such as expressions, gestures, etc., which might be missed out by any audio recorder. Nevertheless, it is still the latter approach that has been chosen for all interviews of this study. Given that the aim of the interview in the study was to collect opinions, the words uttered by teachers were the focus of data collection and data analyses for this method. In addition, the operation of an audio-recording device is simpler than a camcorder which has more steps to follow: choosing an optimal position, setting up the tripod, switching the power on, adjusting the focus, switching the recording button on, … switching the recording button off. Audio recording generally just need three steps: power on, start, and stop. Moreover, to talk with an unfamiliar researcher is generally an unusual task for a teacher, so the use of a camcorder may make the teacher less natural than if s/he is delivering a lesson. In order to leave fewer procedures to the interviewer, and to make the interviewee more relaxed, all interviews were recorded in the form of audio. In case possible recording failures in fieldwork, double recording was applied, with two audio recording apps, installed respectively on a smart phone and a laptop, running simultaneously throughout each interview.

The structure of the interview

The previously mentioned interview framework consisted of eight questions which were clustered around three concepts ranging from (1) teacher beliefs to (2) teacher self-evaluations and (3) teachers' international vision and attitude to practice reform. The purpose of asking the first and second sets of questions was to understand the connections of teachers' beliefs and their self-evaluations. The second question set also served as one of multiple perspectives on the observed lesson. The third question set facilitated individual teachers to foresee mathematics teaching of their own and others horizontally across nations and vertically through time. Moreover, it collected data to complement with focus group data for comparing teachers' perspectives on mathematics teaching in the other participating country before and after actually watching it in focus groups. Detailed questions are as follows:

(1) The teacher's general beliefs of effective teaching was concept one, which contained questions on: *itvQ1* teacher's personal beliefs/descriptions of an effective mathematics lesson, *itvQ2* teacher's strategies for organising and managing lessons, and *itvQ3* teacher's strategies for differentiation so as to meet various learning needs.

(2) The teacher's own reflection and evaluation on the lesson was concept two, which consisted of three questions sequentially about: *itvQ4* teacher's teaching plan, *itvQ5* teacher's reflection on how the plan went, and *itvQ6* teacher's

comments on the lesson and its effectiveness while and after watching the lesson video.

(3) The last concept was about <u>the teacher's international awareness and flexibility</u> which led to two questions: *itvQ7* teacher's imagination of a mathematics lesson to the 9- to 10-year-olds in the other participating country (i.e. England or China), and *itvQ8* teacher's possibility of changing their teaching beliefs and practices in the future.

Data analysis method

As discussed in Section 3.4.4 the analysis of interview data also followed two ways that Cohen et al. (2007, pp. 467-468) suggested, i.e. to organise data by groups and by research question. Teachers' views were first of all netted within each country and then grouped within each interview question. International similarities and differences were discussed where necessary.

3.4.6 Focus groups with teachers (M6)

The focus group method is a way of gathering information from a group of people (usually 5-10) with "certain characteristics in common" and relevant to a specific topic which they will focus on in their group discussion (Krueger & Casey, 2009, p. 2). The formation of focus group interviewing was triggered by the arguments on the limitation of interviewing techniques among social researchers in the 1930s.

The main argument was around the negative effect of interviewers' dominance in the process on the accuracy of interviewing results, as Stuart A. Rice argued:

> A defect of the interview for the purpose of fact-finding in scientific research, then, is that the questioner takes the lead. That is, the subject plays a more or less passive role. Information or points of the view of the highest value may not be disclosed because the direction given the interview by the questioner leads away from them. In short, data obtained from an interview are as likely to embody the preconceived ideas of the interviewer as the attitudes of the subject interviewed.
>
> (Rice, 1931, p. 561)

Following this trend in social research, the focus group interviewing was born in World War II when social scientists started to conduct unstructured interviews in *groups* (e.g., Merton & Kendall, 1946). Though both intended to collect opinions from human participants, individual interviewing and focus group interviewing have a significant difference in that the former emphasises the value of individuals' views, whereas the latter appreciates the collective views which emerge during the process of interactions within the group (Morgan, 1988).

The focus group approach was initially not welcomed by academia, but it was overwhelmingly embraced by the marketing research sector from the 1950s onwards (Krueger & Casey, 2009). The successful application of focus groups in marketing research drew the attention from academics for the second time in the 1980s when academic researchers started to learn from the market people and adjust focus group interviewing for the purpose of academic research. Focus groups can now be conducted either face to face or virtually on the Internet (Lichtman, 2010).

Focus group participants, initially from various natural settings, are temporarily reorganised as new groups to discuss specific topics. Focus groups bring together various experiences and voices on the focused topic(s), and facilitates participants to interact with each other. It provides both processes and outcomes in itself, and sheds light on how a topic is perceived by members of a community as it is in real world. As Morgan (1988, p. 12) remarks, "the hallmark of focus groups is the explicit use of the group interaction to produce data and insights that would be less accessible without the interaction found in a group".

In the focus group, the researcher's role is different from that in an individual interview. The opportunity to discuss in focus groups gives participants full freedom in investigating the topic collaboratively with their peers rather than merely with the interviewer (Hennink, 2007). This decreases the researcher's dominance over the discussion. The roles of participants and the interviewer therefore shifts in ways that 'the group participants take over some of the "interviewing" role, and the researcher is at times more in the position of listening in' (Ritchie & Lewis, 2003, p. 171). Therefore, in a focus group, the researcher is more a facilitator than an interviewer.

It is worth noticing that focus group interviewing usually produces less data than the one-on-one interviewing could possibly do with the same number of participants (Morgan, 1988). Therefore, cautions should be taken before applying focus groups as a data-collection method. The rationale and purpose of using this method should be clarified beforehand and the researcher should demonstrate efficient management skills during the process.

The application of the method in the study

In this study, focus groups were set up for collecting teachers' views on a native and a foreign mathematics lesson to the 9- to10-year-olds. Within each country, teacher participants were divided into two groups. In total, four different groups, EN-FG1, EN-FG2, CN-FG1 and CN-FG2, were organised to watch and comment on the two lessons. There were five teachers in each of English groups; there were initially six in each Chinese group but finally only nine teachers' data were kept for use because three teachers in CN-FG2 (see Section 3.3.3 for detail) dropped out during the testing stage.

Unstructured focus groups

From all lessons observed and recorded in each country, a lesson was chosen for focus group discussions. The chosen lesson should carry rich local features – in other words – it should represent the way local teachers teach. Moreover, the teacher should have a medium length of teaching experience. As will be shown in Table 4.8, 3-5 and 6-10 years would meet this criterion. After taking these into consideration, Lessons EN2 and CN7 were considered as representative of local features. Teacher EN2 were in her fourth year of teaching, and Teacher CN7 had six years' experience. In the focus group, videos of the two lessons, captured by Cam #3, i.e. the main camcorder, were presented sequentially with each followed by a group discussion (see the flowchart shown in Figure 3.5). As in the interview, lesson videos and the effectiveness of the lessons were the foci of the focus group conversation. Apart from the basic features that distinguish individual interviewing and focus group interviewing from each other, in this study, another difference between the two was that, unlike those semi-structured interviews, group discussions were not structured by any specific questions. During the process, topics and topic-related questions arose naturally from the conversation between the participants and were tackled by the participants as well.

The role of the researcher

The unstructured feature of the focus groups in this study significantly minimised the dominance of the interviewer, which increased the authenticity of participants' views collected. Nonetheless, it is slightly different from what Morgan (1988) called "self-managed group", since the researcher is at times to encourage teachers to express their opinions when some of them are too shy to talk (this happened more in China) or when the researcher feels that there are some points they probably missed out. However, the action of the researcher has to stop there. Otherwise, it may significantly reduce the accuracy of data collected.

Relations between focus groups and interviews in this study

Though both focused on the effectiveness of mathematics teaching, focus groups were irreplaceable supplements to post-class interviews. Individual interviews collected teachers' self-perceptions of their practices, which was more of a combination of post-class reflection and self-evaluation. By putting all interview data together, it might also form a bigger picture of how these teachers see the bunch of lessons. Nonetheless, the islands of individuals' views are apparently isolated from one another. This is where focus groups come into play to bridge those *islands*.

The flowchart of the focus group

A discussion guide is regarded as the main data-collection tool for focus group interviewing (Hennink, 2007, Ch.3). It is commonly seen in two forms: a topic guide or a questioning route, with the former consisting of a list of topics or key words and the latter a list of questions (Krueger, 1998).

In the EMT project, the flowchart (Figure 3.5) for the focus group functioned as a discussion guide around the central theme – *the effectiveness of mathematics teaching*. This was consistent from group to group, which made data comparable between groups and countries. Before focus groups, both lessons were translated into another language, either British English or Simplified Chinese, to suit respective contexts. Then, the translations were typed into corresponding videos as subtitles using the apple-based software iMovie in which subtitles were edited and fit accurately in specific time points along the timeline. Then, subtitles in the participants' native language would run simultaneously while the video of a foreign lesson was shown, which made verbal interactions in the foreign lesson understandable to teachers in a focus group from another country.

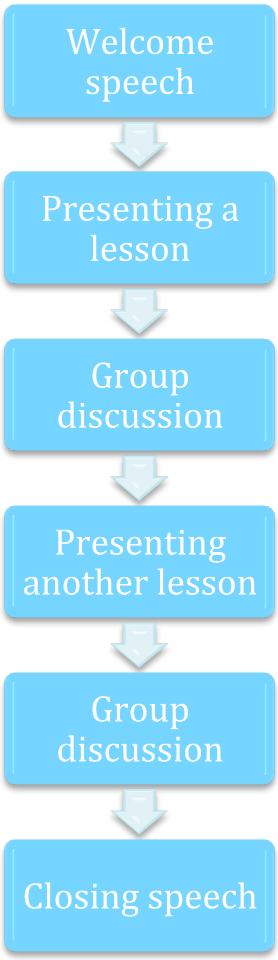


Figure 3.5 The flowchart for focus groups in the EMT project

The approach to recording the focus group

In addition to how to carry out the focus group, another crucial element in the process was how to record the group discussion – note-taking, video-recording or audio-recording (Hennink, 2007, Ch.10)? Taking notes on key points of the conversation are a must-do either for individual interviews or focus groups, but it is digital recordings that work for collecting data with full details. As explained in the section for interviewing, though a video camera is able to capture participants' body languages and facial expressions more accurately than note-taking and audio-recording approaches do, the appearance of a video camera on a tripod may easily intrude into the atmosphere of group discussions and therefore might affect participant spontaneity (ibid.). In addition, the purpose of focus groups for the project is to collect verbal comments on the focused lessons, with which audio recordings could serve well. Hence, audio recording was also applied to the focus groups using the same recording facilities as those used for the interviews.

International focus groups

There are precautions for the researcher to take, when it comes to international focus groups (Krueger & Casey, 2009, Ch.10). Efforts need to be made so as to pose little pressure to English teachers from the researcher who was both a foreigner and an outsider to them. Using a local language may considerably shorten the distance between the insider and the outsider (Krueger & Casey, 2009, p. 171). Timing was also an essential issue that demands careful consideration. Focus groups were all scheduled at a later stage than other events in each site, generally before the final step of data collection – the mathematics test, when teachers were more familiar with the researcher in the process of taking part in prior events.

Data analysis method

The interpretation of focus group data followed the similar approach to analysing interview data as suggested by Cohen et al. (2007, pp. 467-468). Data were organised by groups (nations) and issues (lessons).

3.5 Ethical issues

Children are relatively weaker human beings both physically and emotionally, so they need more protection particularly when it comes to participating in a research project (Punch, 2009, Ch.3). Primary pupils in both countries are regarded as immature citizens whose participation consent forms need to be signed by their parents/guardians. The way in which the study is described is crucial to making parents/guardians aware of the safety level of the study and in obtaining their consents for their children's participation. For this study, during the data collection process, precautions were taken in that videoing facilities should be placed out of the way, to ensure children could walk and act normally and safely. The researcher acted politely always with a smiling face.

Video-based studies might expose risks on "the confidentiality and private rights of those individuals who are filmed" (Jacobs et al., 2007, p. 290). Therefore, ethics is a sensitive issue in this study. It was a relatively complex process to acquire consents to use recorded information from participants for research purpose, because this involves multi-level approvals, mainly from parents/guardians, school leaders, and teacher participants.

Because of differences in cultures and conventions in the two countries, plans for every step of data collection need to be adjusted accordingly in different contexts. Always following the way of the native people is the top principle in such adjustment.

Ethical application documents, including consent forms and information sheets for teachers and parents/guardians, were drafted by the research student, and discussed at supervision meetings. Meanwhile, data-collection instruments were also decided and ready for official checks. At the preparation stage, the main considerations were on the clarity of research aims and procedures explained in these documents, the protection of data proposed to collect from participants, and the acceptability of all data-collection events to various participants on both sites.

After all revisions had been done, the ethical clearance application packet, mainly containing data-collection instruments, consent forms and information sheets, was submitted via the University's online ethical system, Ethics and Research Governance Online (ERGO). At the requirement of the ethical committee of the University, minor amendments were made twice, and then the ethical clearance of the EMT project was confirmed (as shown in Appendix A). With the approved consent forms, information sheets and data-collection instruments, the researcher started to approach and recruit participants in both countries.

Throughout the data collection and analyses, cautions were given to every aspect foreseen in the preparation stage and arising in the actual researching processes. Particular attention was given to the role of the researcher striving to be as impartial as possible in and across the contexts of two different cultures. A strong commitment to the Education science enthused the researcher to plan, act, reflect upon and report the study in a position as neutral as possible where she perceived herself as a world citizen rather than a Chinese. The supervision offered by a British Supervisor in a British institution – including activities, such as the inter-rater reliability test on observational data – consistently reminded the Chinese PhD student of her obligation in taking an ethically neutral stance whilst maintaining self-conscious on the research journey (Weber, 1949).

In a classroom setting, a researcher would in each context mean a little bit of an outsider to the 'original inhabitants' there. One of the efforts that the investigator made was to approach teachers in advance and familiar herself with them to reassure them that she shared the same passion with them in the pursuit of a better education for every child. Approximately five minutes before the data collection of each event, she would ask the teacher to allow her to introduce or reintroduce herself to pupils in the class, remind them of what had been explained in the research information sheet and tell them the purpose of the specific event in 'their language'. By doing so, she was able to calm their nerves, if anyone felt nervous because of the presence of a (foreign, in the English case) researcher. The researcher anticipated that such an effort in England might need to be bigger than in China, and she came prepared with this in mind, but in reality English pupils turned out to be emotionally indifferent from their peers in China – they all behaved in a natural state.

In the process of data analyses, the researcher carried on holding a deep interest in mathematics teaching and learning. She often reminded herself that, on this journey, a blind passion for any nation or culture or anything other than the advancement of maths teaching and learning sciences was an absolute enemy that must be defeated. She also alerted herself to her own background and experience and tried her best to *not* let these affect her neutral interpretation of the data and initial findings.

Other details about the ethical efforts that the researcher has made are also discussed in former sections on specific research methods.

3.6 Chapter conclusion

This chapter clarified the philosophical position of the study, located research methods and research questions in the theoretical framework, and explained the methodological details of the study by linking research questions to research methods. Parallel to that, attention has also been drawn to essential methodological work and methodological designs of prior studies in similar or partially similar areas. Thus, this chapter helped interweave the research design of this study with the intellectual heritage succeeded from the literature together. This in turn may make potential contribution to the methodological evolution of teacher effectiveness research in the subject of mathematics in the global context.

The application of multiple methods in addressing research questions brings together scientific evaluation of teaching and learning and multiple perspectives from various roles in and beyond the classroom. This in turn may contribute to future collaborations between educational effectiveness researchers and educational improvement researchers with their methods mutually supplemented in one study. The methodological emphasis in this study lies on the interconnections between multi-layered data and findings generated by these methods. Hence, the design of the study is purpose-driven, instead of paradigm-driven.

Results and findings are to be reported in the following two chapters. Chapter 4 focuses on the results of quantitative analysis of data; Chapter 5 gives detailed interpretation of data analysed in a qualitative way.

CHAPTER 4 - RESULTS & FINDINGS I

Chapter overview

- ✧ *Chapter introduction*
- ✧ Structured lesson observations (M1)
- ✧ Teacher questionnaire (M2.1)
- ✧ Pupil questionnaire (M2.2)
- ✧ Standardised mathematics tests (M3)
- ✧ Correlations between maths teaching & learning
- ✧ Pupil background differences
- ✧ *Chapter conclusion*

4.1 Chapter introduction

This chapter provides detailed results of analyses of quantitative data from structured observations, teacher questionnaire, pupil questionnaire, two standardised tests and correlations between teaching and learning outcomes.

4.2 Structured lesson observations (M1)

To systematically understand the quality of teaching, lessons were each measured against the two observation systems – the Opportunity to Learn (OTL) and the International System for Teacher Observation and Feedback (ISTOF).

4.2.1 Measuring the quantity of teaching with OTL

As explained in Chapter 3, for each teacher's lesson, two types of time-related percentages were calculated to show time on five categories of classroom activities and pupil time on task.

Figure 4.1 gives an overview of the duration of lessons across countries. Chinese lessons ranged from 36 to 45 minutes, whereas English lessons lasted longer, ranging from 47 to 66 minutes.

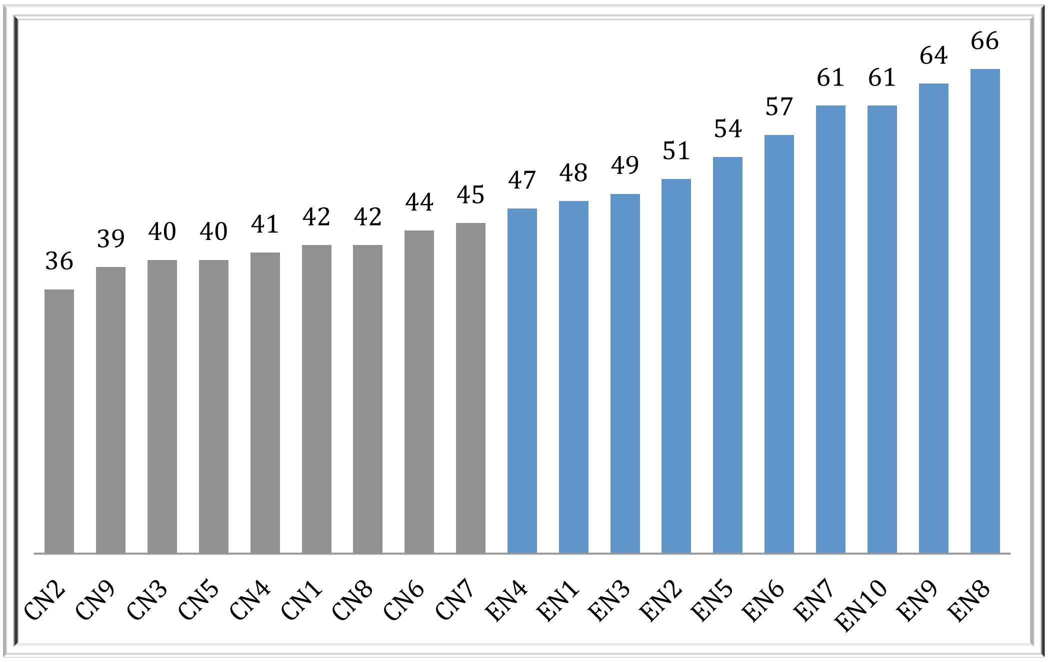


Figure 4.1 Duration of lesson time across classrooms

The following sections will present two strands of OTL results: (1) classroom activities and (2) pupil time on task.

4.2.1.1 OTL 1-5: Five types of classroom activities

This section gives detailed descriptions of the distribution of lesson time to five types of OTL activities across classrooms and countries:

- Whole class interaction (OTL1)
- Whole class lecture (OTL2)
- Individual/group work (OTL3)
- Classroom management (OTL4)
- Partial class interaction (OTL5)

OTL 1 – Whole class interaction

Whole class interaction happens when the teacher interact with pupils in the whole class through questioning or discussion. It is different from another OTL activity, *whole class lecture*. The distinction between the two is whether there is one-way or two-way communication between the teacher and pupils.

Table 4.1 Number of whole class interaction segments in lessons

EN1	***EN2***	***EN3***	***EN4***	***EN5***	***EN6***	***EN7***	***EN8***	***EN9***	***EN10***
3	1	6	3	9	3	5	3	3	3
CN1	***CN2***	***CN3***	***CN4***	***CN5***	***CN6***	***CN7***	***CN8***	***CN9***	
7	6	6	10	5	10	7	7	6	

As shown in Table 4.1, the number of *whole-class interactive* activities varied significantly across countries. Chinese teachers not only spent more percentages of lesson time on *whole class interaction*, but also tended to split the time into more chunks. Roughly, the numbers of whole-class interaction segments in Chinese lessons doubled those in English lessons, except for Teacher EN3 and Teacher EN5's lessons.

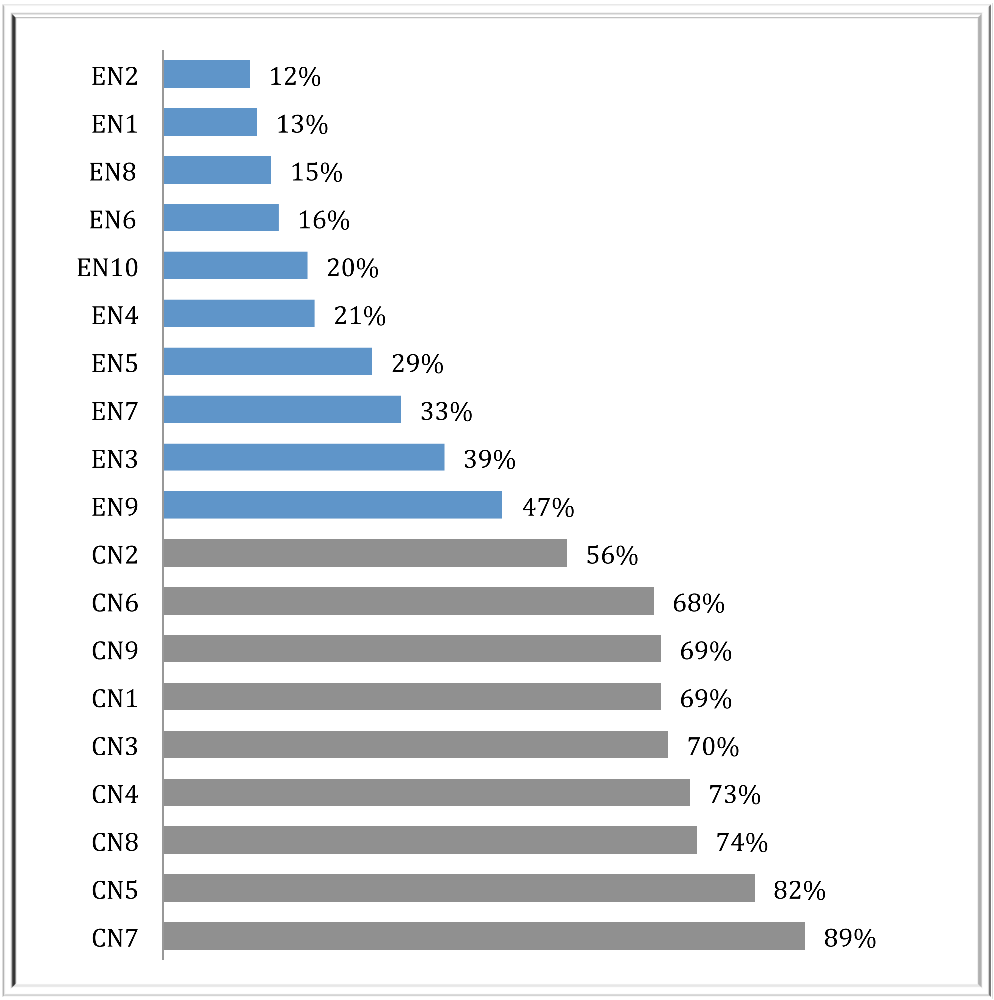


Figure 4.2 OTL 1 – Lesson time on whole class interaction

Proportionally, it is apparent in Figure 4.2 that all Chinese teachers spent more time on *whole class interaction* than did English teachers. The frequency analysis aided by SPSS provides more details. The corresponding percentage in 10 English classes ranges from 12% to 47%, which makes a mean of 23.8% and a standard deviation of 11.6%; the percentage in 9 Chinese classes ranges from 56% to 89%, with a mean of 72.2% and a standard deviation of 9.3%. Across all nineteen participating classes, the average percentage of time on *whole class interaction* is 46.7%, and the standard deviation 26.9%.

It is obvious that intra-country differences are smaller than the inter-country differences (t (17) = 9.97, p = < .001), which once again indicates the importance of conducting cross-national TER studies – to see the full picture. The effect of nationality on the proportion of time allocated for whole class interaction is strong (d = 4.63 > 1).

OTL 2 – Whole class lecture

Whole class lecture happens when the teacher is delivering information directly to the class without pupils' active participation and response. Table 4.2 shows the numbers of *whole class lecture* segments in all classrooms. It is apparent that seven England's lessons had such activities.

Table 4.2 Number of whole class lecture segments in lessons

EN1	*EN2*	*EN3*	*EN4*	*EN5*	*EN6*	*EN7*	*EN8*	*EN9*	*EN10*
1	2	1	2	4	1	0	1	0	0
CN1	*CN2*	*CN3*	*CN4*	*CN5*	*CN6*	*CN7*	*CN8*	*CN9*	
0	0	0	0	0	0	0	0	0	

The striking finding is that Chinese teachers did not lecture to the whole class at all across the nine Chinese classes, whereas seven out of ten English teachers chose to do so during a period of 3% to 9% of their lesson time (Figure 4.3). England's average percentage for *whole class lecture* is 3.9%, with the standard deviation being 3.2%.

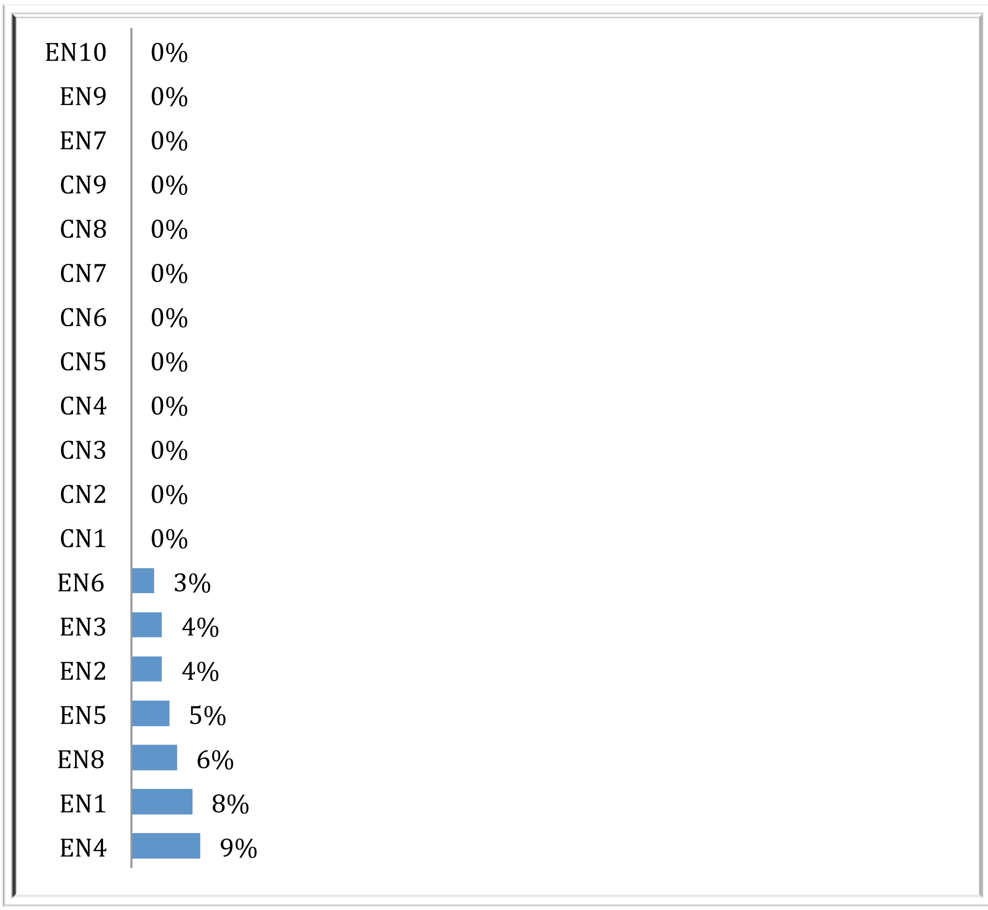


Figure 4.3 OTL 2 – Lesson time on whole class lecture

OTL 3 – Individual/group work

How often a teacher used *individual/group work* also varied a lot both within and across countries. Table 4.3 shows the number of segments coded into this type of activity in each lesson.

Table 4.3 Number of individual/group work segments in lessons

EN1	*EN2*	*EN3*	*EN4*	*EN5*	*EN6*	*EN7*	*EN8*	*EN9*	*EN10*
7	3	6	5	11	4	8	3	2	7
CN1	*CN2*	*CN3*	*CN4*	*CN5*	*CN6*	*CN7*	*CN8*	*CN9*	
5	6	5	9	4	9	6	6	6	

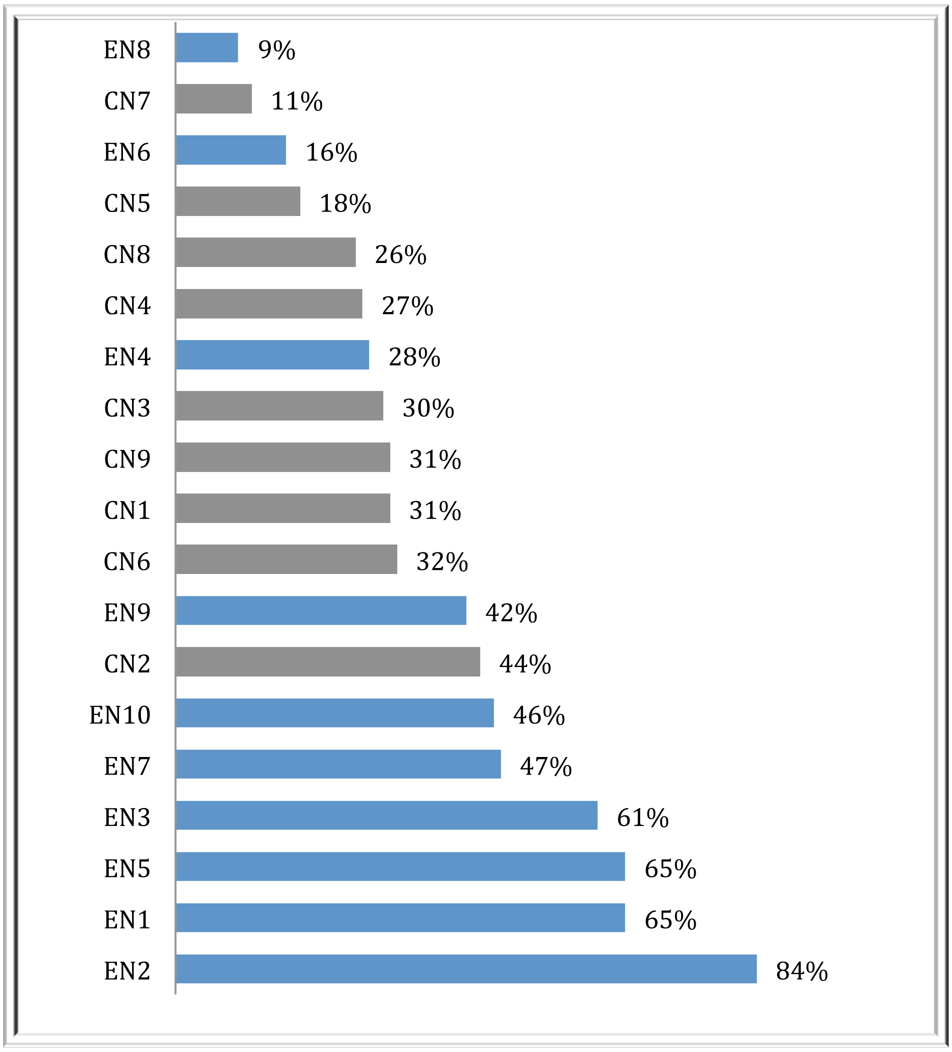


Figure 4.4 OTL 3 – Lesson time on individual/group work

Moreover, though all teachers in both countries have arranged time for pupils to work individually or in pairs mainly on written tasks, the percentages of lesson time for this type of activity varied significantly both within England and between two countries (see Figure 4.4). The international variation can also be reflected by the means and standard deviations calculated at the national and international levels: England has a mean of 46.6% (SD = 24%); China has an average of 27.8% (SD = 9.3%); pooling both countries, the international mean is 37.7% (SD = 20.4%). The international difference is statistically significant (t (12) = -2.3, p = .04 < .05) and the nationality of each class has a strong effect on its proportion of lesson time on independent/group work (d = 1.13 > 1).

OTL 4 – Classroom management

Classroom management does not directly involve subject matter teaching, and it often occurs when the teacher is organising pupils and/or other materials for either activity preparation or discipline purposes. In the EMT project, eight out of ten English teachers had spent time on this type of activity; none of Chinese teachers had done so (Table 4.4).

This type of activities occupied on average 3.5% (SD = 2.4%) of English lesson time (Figure 4.5). The discipline of Chinese classes was better than that of English classes, and the transition between activities were also speedy and natural in Chinese classes. It thus seemed unnecessary for Chinese teachers to 'manage' activities.

Table 4.4 Number of classroom management segments in lessons

EN1	*EN2*	*EN3*	*EN4*	*EN5*	*EN6*	*EN7*	*EN8*	*EN9*	*EN10*
3	0	0	1	1	2	2	2	3	3
CN1	***CN2***	***CN3***	***CN4***	***CN5***	***CN6***	***CN7***	***CN8***	***CN9***	
0	0	0	0	0	0	0	0	0	

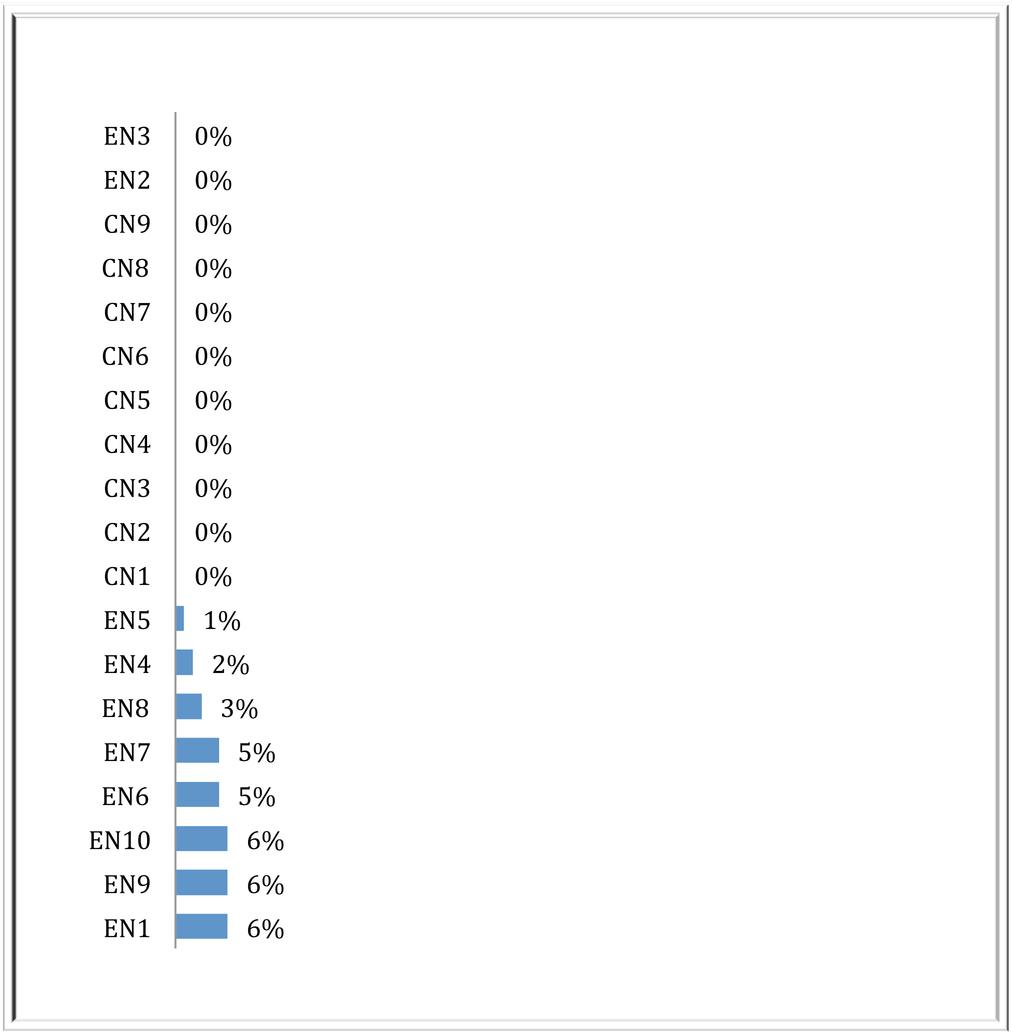


Figure 4.5 OTL 4 – Lesson time on classroom management

Another type of OTL classroom activities is partial class interaction the results of which will be presented as follows.

OTL 5 – Partial class interaction

Partial class interaction is the type of activities in which the teacher is conducting two-way interactions with individuals or part of the class rather than the whole class. This was a typical phenomenon only found in the English classes where setting by ability was commonly practised; this type of interaction did not happen in Chinese classes (Table 4.5 & Figure 4.6).

Table 4.5 Number of partial class interaction segments in lessons

EN1	***EN2***	***EN3***	***EN4***	***EN5***	***EN6***	***EN7***	***EN8***	***EN9***	***EN10***
3	0	0	1	0	5	4	2	1	4
CN1	***CN2***	***CN3***	***CN4***	***CN5***	***CN6***	***CN7***	***CN8***	***CN9***	
0	0	0	0	0	0	0	0	0	

In the English classroom, this type of activities often happened on the carpet in front of the IWB. In most cases, the teacher's purpose of interacting with the partial class was to re-teach low-ability pupils the content that had been taught by the teacher but not well understood by some pupils; often during the partial class interaction time, the rest of the class were working independently on worksheets.

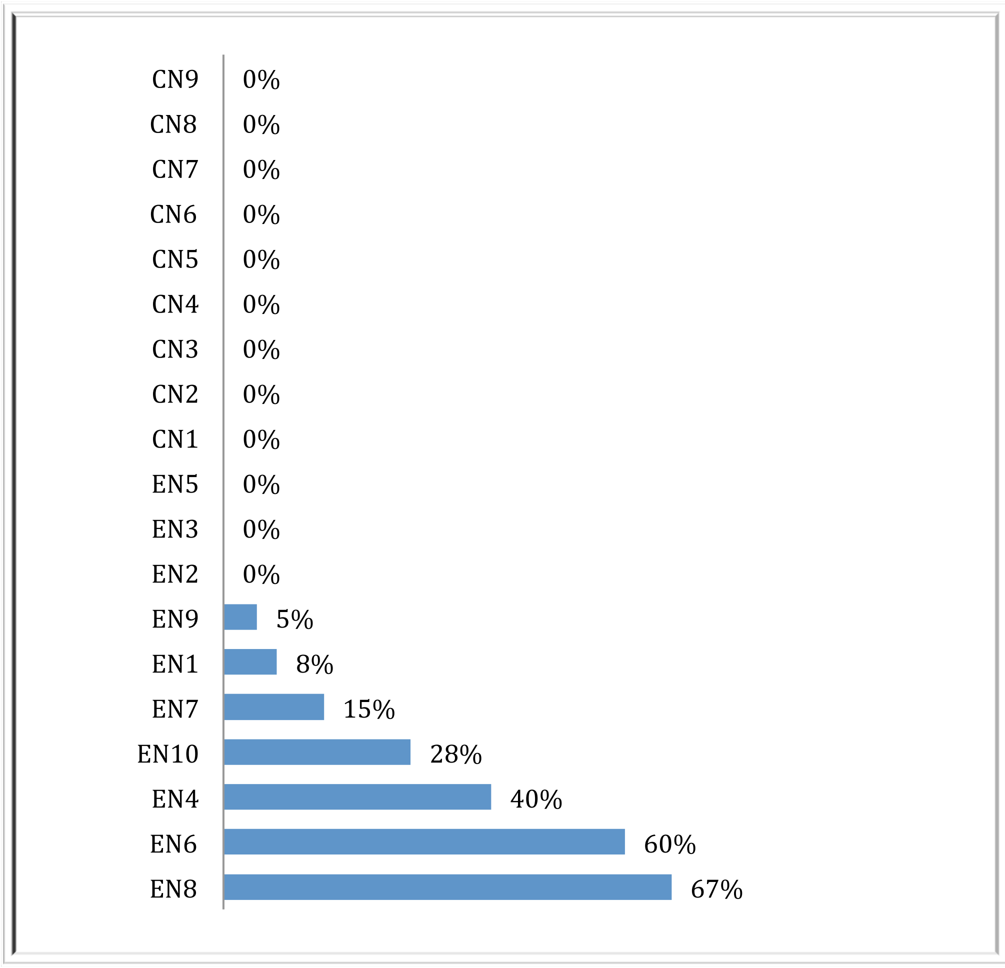


Figure 4.6 OTL 5 – Lesson time on partial class interaction

Most of the time, English teachers did not ask pupils to come forward to the carpet; pupils actively chose to come whenever they felt frustrated. In very rare cases, the teacher's purpose was to teach more advanced content to high-ability pupils. Overall, in English classes, the average percentage of lesson time on partial class interaction is 22.3% (*SD* = 25.4%) of lesson time.

Now that we have had a picture of how lesson time was distributed into five different types of classroom activities in each of the nineteen classrooms, we will carry on to see the percentages of pupils on task during even intervals of the lesson time.

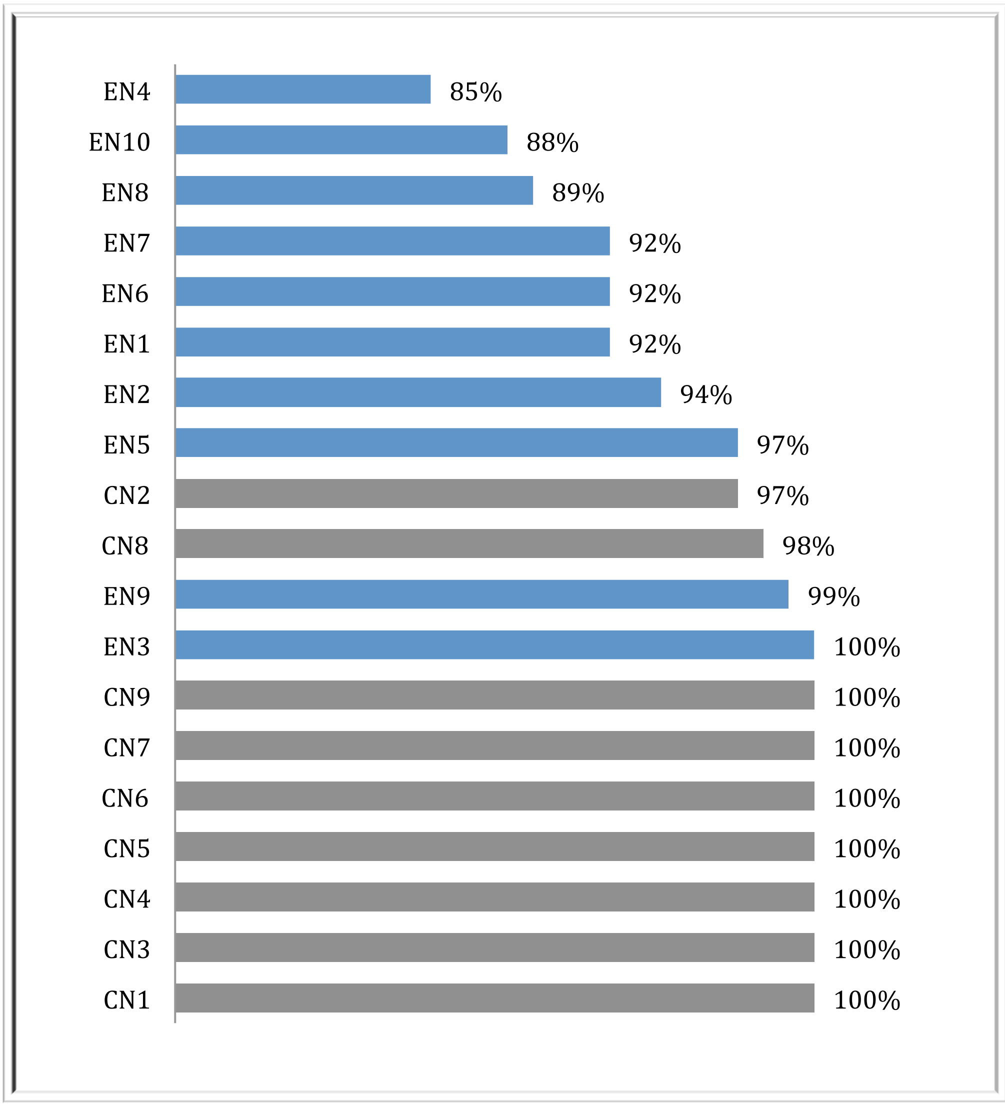


Figure 4.7 OTL 6 – Pupil time on task across classrooms

4.2.1.2 OTL 6 - Pupil time on task

The result shows that English classes in general have lower time on task than do Chinese classes. In Figure 4.7, of ten English classes, one has all pupils on task throughout of the lesson, whereas seven out of nine Chinese classes have no pupils off task throughout of the lessons. The average percentage of pupils on task within England was 92.8% (SD = 4.8%), and the national average for China was 99.7% (SD = 0.7%), with the international mean being 96.1% (SD = 4.9%). The relationship between nationality and pupil time on task was statistically significant with a strong effect (t (9) = 4.5, p = .001, d = 2.5 > 1).

4.2.1.3 Summary on the OTL measure

OTL measured the quantity of teaching factors within the frame of lesson time. Overall, as shown in Figure 4.8, English teachers allocated almost half of the lesson time for pupils working individually or with their peers on exercises, whereas, in about three quarters of their lesson time, Chinese teachers were interacting with the whole class.

English teachers allocated a varying proportion of lesson time to each of the five types of activities, whereas in all Chinese lessons, there were only two types: about three quarters of whole-class interaction and about a quarter of individual/group work.

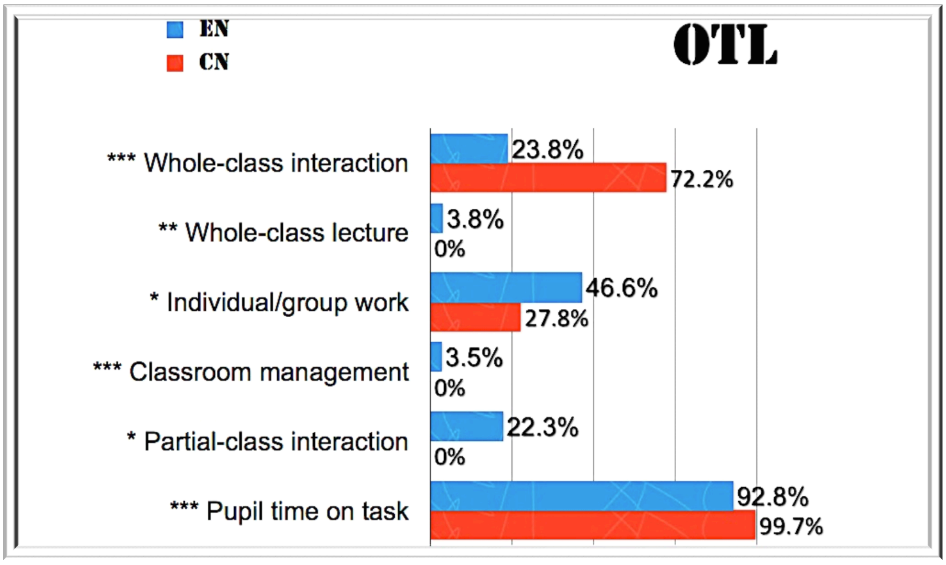


Figure 4.8 The international comparison of six OTL percentages

On average, Chinese classes had a higher proportion of pupils on task, in comparison with English classes. In fact, only one pupil in each of two Chinese classrooms was off task over a short period, whereas only one English class had 100% pupils on task throughout the lesson. Chinese teachers did not appear to take actions to manage the class but were able to keep children on task; English teachers used an average of 3.5% of lesson time to manage the class but had more children off task. These results, representing the *quantity* of teaching, are correlated with pupil test scores in certain ways, which may be found in section 4.6.4. Next comes the measure of the *quality* of teaching.

4.2.2 Measuring the quality of teaching with ISTOF

The following sections will present the results of teacher scores (i.e. quality of teaching) in the six domains of ISTOF teacher behaviours (*N* = 40, the EMT adapted version) which are clustered around 19 behavioural indicators (in brackets):

- ISTOF 1 – Assessment and evaluation (ISTOF 1.1 & 1.2)
- ISTOF 2 – Clarity of instruction (ISTOF 2.1, 2.2 & 2.3)
- ISTOF 3 – Instructional skills (ISTOF 2.1, 3.2 & 3.3)
- ISTOF 4 – Promoting AL & MSs[2] (ISTOF 4.1, 4.2, 4.3 & 4.4)
- ISTOF 5 – Classroom climate (ISTOF 5.1, 5.2, 5.3 & 5.4)
- ISTOF 6 – Classroom management (ISTOF 6.1, 6.2 & 6.3)

Within each domain, firstly a summary of the results will be presented, and secondly teacher behaviours in two countries' classrooms will be described as being observed and measured against each behavioural indicator.

4.2.2.1 ISTOF 1 – Assessment and evaluation

This component contains 4 scales, each ranging from 1 (strongly disagree) to 5 (strongly agree), which makes the full score of this domain 20. The average score of England is 12.8 (*SD* = 2.3), and that of China is 19.4 (*SD* = 1), with the international mean score being 16 (*SD* = 3.8) (Figure 4.9). The international

[2] active learning and metacognitive skills

difference is statistically significant (t (17) = 8, $p < .001$) with a strong effect (d = 4 > 1).

Teacher behaviours in this component are clustered around two indicators, ISTOF 1.1 and ISTOF 1.2.

ISTOF 1.1 – Did the teacher give explicit, detailed and constructive feedback?

Regarding answers or solutions to given problems, English teachers tended to state what was right rather than further scrutinise why it was right. Chinese teachers tended to facilitate pupils themselves to think and find the answers and/or a variety of solutions if possible and also promote them to think why some pupils made it correctly whereas others did it incorrectly and how to avoid similar mistakes afterwards. They tended to prepare and carry out carefully the teaching arrangement between posing a question and achieving an agreed answer/solution, so that pupils' cognition was developed step by step towards realising the answer/solution themselves. Therefore, a clear distinction between English and Chinese teachers' approaches to questioning was who ultimately found the answer and told the truth – pupils or the teacher?

In responding to pupils' answers, English teachers' feedback was often nice and simple, and they seemed to hold the position in telling the truth and making formal judgement in the class; Chinese teachers tended to ask the rest of the class to judge whether a pupil's answer was correct or wrong; they were more likely to follow a pupil's answer with another question – often deeper than the prior question and at the same time building on the prior answer.

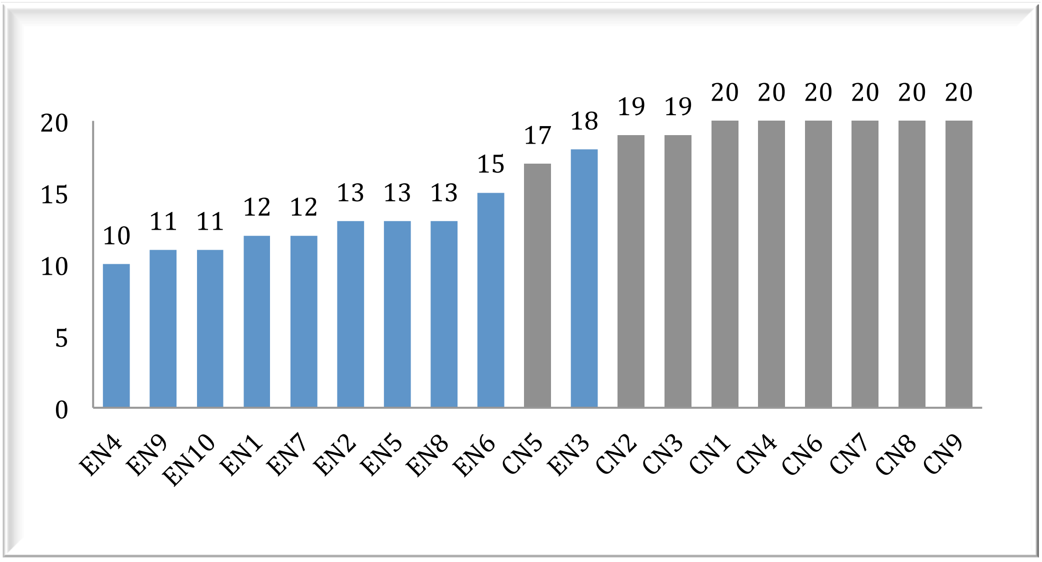


Figure 4.9 ISTOF 1 – Assessment and evaluation

ISTOF 1.2 – Was assessment aligned with goals and objectives? All teachers from both countries had provided their pupils' assignments that were clearly related to what they learned. However, English teachers tended to stop, once learning objectives introduced and related assignments given; Chinese teachers focused more on the way in which assignments were connected with the learning goals.

Next, we will look at the second ISTOF domain – *clarity of instruction* which carries three indicators and six scales.

4.2.2.2 ISTOF 2 – Clarity of instruction

This component contained six items with a total score of 30. The English average is 16.4 (*SD* = 4.4), the Chinese average 27.9 (*SD* = 1.6), and the international mean 21.8 (*SD* = 6.7).

Figure 4.10 shows that most Chinese teachers outscored their English counterparts. Statistically, the international difference is significant, and the relationship between nationality and teacher scores in this domain is strong (t (12) = 7.7, p < .001, d = 3.8 > 1). Detailed explanations are as follows.

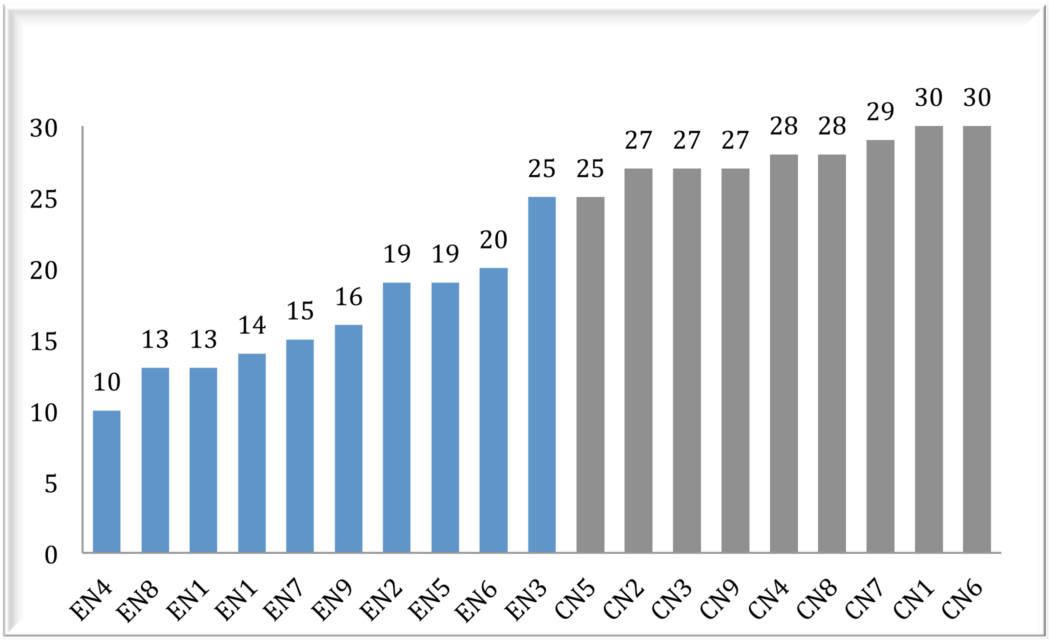


Figure 4.10 ISTOF 2 – Clarity of instruction

ISTOF 2.1 – Did the teacher show good communication skills? English teachers sometimes checked for understanding, but other times they simply carried on and dragged pupils' mind through the procedures they pre-defined for solving a certain type of problems, no matter pupils got it or not. In most classrooms, pupils did not get it at first, and the teacher had to re-teach the content for multiple times. Chinese teachers asked lots of questions and sought their awareness of pupils' status of understanding through their answers. Overall, Chinese teachers communicated with the class in a clearer and more understandable manner than did their English colleagues.

ISTOF 2.2 – Did the teacher give clear explanation of purpose? All teachers in both countries clarified the lesson objectives during the first five to ten minutes. There were however some international differences: English teachers all gave the learning objectives – so called *Success Criteria* – at the very beginning of the lesson; Chinese teachers tended to pose real life problems/questions and involve pupils into a whole-class discussion which eventually yet smoothly led the class to the learning topic. In addition, there were more reasoning activities going on in the Chinese classroom than in the English. Chinese pupils got more opportunities to think actively the interconnections between the importance of learning targets, the relevant real world issues, their prior knowledge and the given tasks; English pupils had more chances to think relatively narrowly that everything they did was mainly (if not solely) for the purpose of reaching the steps in their *Success Criteria*.

ISTOF 2.3 – Were lessons well structured? English teachers only applied relatively less time in formal instruction of new knowledge, and the rest of the time they were mainly re-teaching the same content to individuals or a small group of individuals. During the formal teaching period, some (i.e. not all) English teachers tended to present the lesson content in a logical flow moving from simple to more complex concepts; during the majority of lesson time, most teachers were often busy re-teaching what they had just taught strictly following the same procedure as for the first input; the transition between lesson components was rough and lacked coherence.

Chinese teachers spent almost two thirds of the lesson time gradually developing children's cognition from old to new knowledge; the strong logical flow was commonly existent in the Chinese classroom; the construction of mathematical knowledge was steady and coherent moving from simple to complicated concepts and/or varying situations; the transition between lesson components was natural and seemingly necessary at each specific time point because everything had been readily built for the transition; the smooth transition kept the momentum of teaching and learning from the beginning throughout the lesson to the end.

Overall, most English lessons were partially well structured during the period of formal input and partially poorly structured during independent work periods; Chinese lessons were well structured throughout the lesson time.

The next domain of ISTOF is *instructional skills* which consists of 3 indicators and 5 behavioural scales.

4.2.2.3 ISTOF 3 – Instructional skills

On the five 5-point scales of this ISTOF domain, English teachers' scores ranged from 7 to 14, and Chinese teachers scored from 20 to 25. The English mean is 11.4 (SD = 5.1), the Chinese mean 23.2 (SD = 1.7), and the international mean 17 (SD = 7.1) (Figure 4.11). Again, the international difference is statistically significant and the impact of nationality on teacher scores in this domain is strong (t (17) = 6.6, p < .001, d = 3.5 > 1).

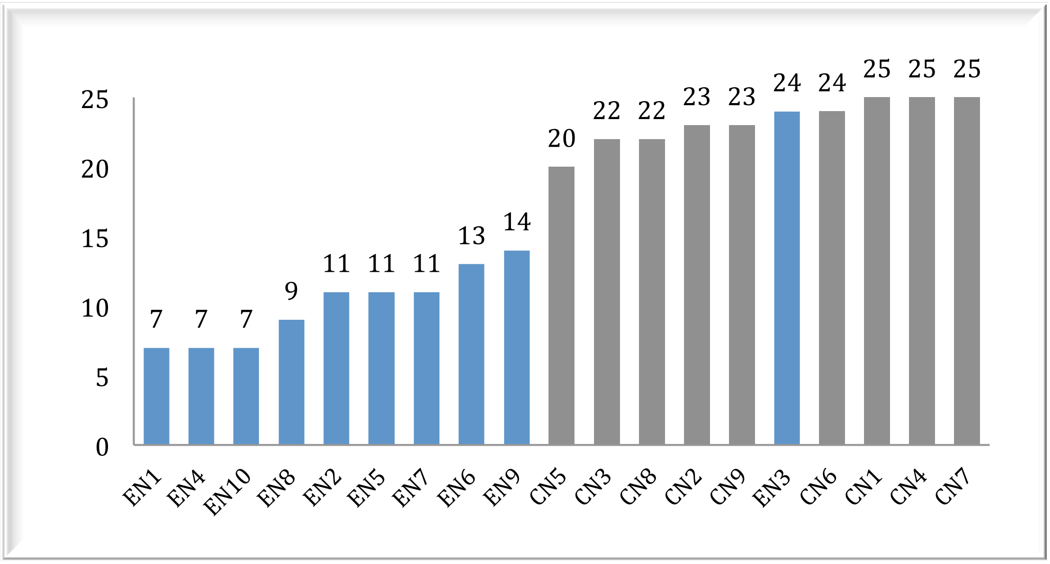


Figure 4.11 ISTOF 3 – Instructional skills

The following paragraphs provide detailed accounts of teaching in two countries regarding the 3 indicators.

ISTOF 3.1 – Was the teacher able to engage pupils? The main methods of engaging children in learning in both countries were teacher questioning and classwork. In terms of questioning, Chinese teachers not only asked more questions but also asked different pupils to offer supplement/alternative answers to a single question, which helped engage pupils more deeply into mathematical thinking; English teachers asked fewer questions in a lesson, and they tended to keep a question to each individual, building little connection between individuals' answers. Furthermore, Chinese teachers were very unlikely to tell the correct answer to a given question, and they tended to let the correct answer emerge from pupils in the wake of whole-class discussion. Nonetheless, in the English lesson, if the correct answer did not come out from the pupil being questioned, the teacher utmost asked two more pupils to have a go. After that, the teacher was ready to tell the correct answer and explain detailed procedures, which did not pose enough challenge to children and allow them to think actively out of their comfort zone.

The assignments in the Chinese classroom were tailored appropriately in terms of difficulty level and timing. The Chinese lesson content was organised along the line of examples in a sequence such that, as time went on, the difficulty level and depth of knowledge increased. Following each example, the teacher would ask pupils to work on a similar task independently. Each independent task was given at the time point when the class had collectively come up with one or multiple solutions to the example. Because of the prior whole-class questioning and answering, pupils were ready to complete the given task quickly (often 1-2 minutes per task) and accurately. The assignments in the English classroom were also tailored appropriately but were strictly in line with the *Success Criteria* rather than with children's cognitive readiness; the assignment time came, in most cases, too early to allow teachers to make the new knowledge thoroughly absorbed by pupils. This also explained why English pupils kept coming to the carpet for repetitive teaching during independent work. They could not apply the new knowledge before they had fully understood and firmly grasped it.

ISTOF 3.2 – Did the teacher possess good questioning skills? English teachers tended to pose product questions which did not require complex reasoning and deep thinking. On the contrary, Chinese teachers asked more questions – mostly process questions which promoted mathematical reasoning and deeper thinking. The length of the pause following each question did vary a bit according to the difficulty level of questions in both countries' classrooms. Nevertheless, Chinese teachers tended to pause for an optimal length, and if the pause did not work, they tended to pass the question on to another pupil that might be able to answer it, hence keeping the momentum of the lesson. English teachers often gave the full explanation of the answers if a correct response had not appeared, which they seemed to believe was part of their duty; on the contrary, when encountering similar situations, as explained in the indicator ISTOF 3.1, instead of providing the answers themselves, Chinese teachers tended to persevere seeking the next pupil to answer the question until one or a range of reasonable answers emerged.

ISTOF 3.3 – Did the teacher use various teaching methods and strategies? English teachers tended to use simple teaching methods and follow fixed strategies which seemed to shine less surprises through pupils' eyes, even though they all had amazing IWBs which could only be found in the classrooms of some commercial after-school courses in China. Chinese teachers, with less advanced equipment however, were able to adjust their teaching methods and arrangements every now and then so that they could keep pupils *awake* and engaged throughout a lesson; it appeared that the change of methods in the Chinese classroom was not in itself for the sake of change, but for the purpose of better developing children's certain aspect of knowledge at that specific point in the lesson.

In the following section, we will move on to look at the evaluative results of teacher behaviours in the 4th ISTOF domain, *promoting active learning and developing metacognitive skills*, which includes 4 indicators and 10 behavioural scales.

4.2.2.4 **ISTOF 4 – Promoting active learning & developing metacognitive skills**

On the ten 5-point scales in this domain, English teachers attained an average of 21.7 (SD = 6.9), whereas Chinese teachers had a mean of 47.3 (SD = 2.9). The international mean was 33.8 (SD = 14.1) (Figure 4.12). There was a statistically significant and strong relationship between nationality and teacher scores in this domain (t (17) = 10.3, p < .001, d = 5.2 > 1).

In general, English teachers tended to ask product questions which led to exact facts or procedures, but Chinese teachers more often asked process questions before and after a task to generate ideas as to how the task might be tackled or had been tackled and why. English pupils had to follow teachers' instruction of the *right* facts or procedures, whereas Chinese pupils had rich opportunities to plan and discuss strategies or reflect upon their and their peers' problem-solving experiences – either successes or failures. In the English class, there was often one set solution/procedure to a problem; in the Chinese class, solutions were often more than one and procedures could be flexible.

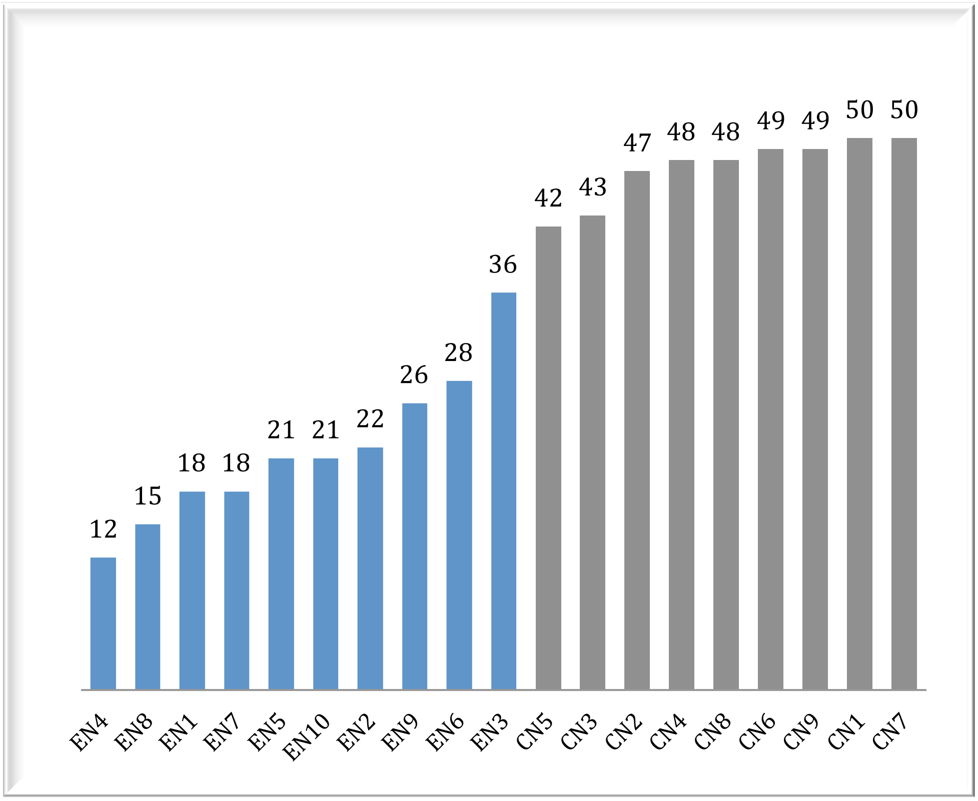


Figure 4.12 ISTOF 4 – Promoting active learning & metacognitive skills

ISTOF 4.1 – Did the teacher help pupils develop problem-solving and metacognitive strategies? English teachers tended to invite pupils to use the strategies that they had taught them; Chinese teachers encouraged pupils to think of multiple strategies to solve a varying problem and share their strategies during whole-class discussion. English teachers were more likely to talk the pupils through the problem-solving steps than asking them independently explain detailed strategies; Chinese teachers tended to ask pupils to explain various solutions they could come up with and foster brainstorming discussions among them. Most teachers from both countries tended to provide instruction in problem-solving processes. Nevertheless, English teachers were likely to deliver the *right* strategies straightaway or after the first attempt of questioning, whereas Chinese teachers tended to ask a series of questions of varying depths to provoke children's thoughts so they could gradually find the strategies by themselves.

ISTOF 4.2 – Did the teacher give pupils opportunities to be active learners? All teachers from both countries encouraged pupils to ask one another questions and to explain their understanding of the topics to one another. English teachers tended to promote paired discussions; Chinese teachers inclined to whole-class discussions. All teachers would encourage pupils to correct their own work if they spotted pupils' mistakes. However, because there was lots of individual work parallel with seemingly constant re-inputs on the carpet in the English classroom, English teachers did not have extra energy to capture all errors emerged in children's classwork, let alone asking them to correct each error. In the Chinese classes, pupils could see clearly what counted as correct and wrong solutions and why during whole-class interaction; during the independent-work period, Chinese teachers all circulated fairly quickly through the class to check *every* pupil's work; they often picked up representative work samples – correct and wrong ones – to share with the class via the projector immediately after each slot, asking pupils to make the judgement, comments and corrections if needed.

ISTOF 4.3 – Did the teacher foster critical thinking in pupils? English teachers focused more on transmitting standard solutions (often just one solution) rather than inspiring diverse solutions; they therefore did not frequently ask pupils to judge the pros and cons of different approaches. It was typical in Chinese classrooms that multiple solutions were always applauded and pupil-led comparisons of different approaches were particularly encouraged. English teachers stopped scrutinising pupils' internal thinking processes once they made sure the correct solutions/procedures were well delivered from their own points of view. Chinese teachers however demanded deeper thinking amongst pupils; every task they gave would include pre- and post-task discussions; the latter nurtured post-action reflection on problem solving, which seemed similar to action research. Despite differences, all teachers from both countries invited pupils to give their personal opinions on certain issues.

ISTOF 4.4 – Did the teacher connect materials to pupils' real world experiences? English teachers tended to discuss mathematical content with pupils for the sake of mathematics itself; Chinese teachers always started and ended their lesson in relevant phenomena in daily life and related the lesson content to the real world whenever necessary throughout the lesson. English pupils were not often invited to raise examples from their life experiences, whereas their Chinese peers were more frequently asked to share personal experiences that were related to the mathematical content throughout the lesson.

The next ISTOF domain is *classroom climate* which has 4 indicators and 8 behavioural scales.

4.2.2.5 **ISTOF 5 – Classroom climate**

The domain of classroom climate contains eight 5-point scales. The international average was 30.3 (*SD* = 9.2), with English teachers averaged 22.8 (*SD* = 6) and Chinese teachers 38.8 (*SD* = 1.1) (Figure 4.13). The international difference was statistically significant, and the relationship between nationality and teacher scores was strong (t (10) = 8.3, p < .001, d = 4.5 > 1).

ISTOF 5.1 – Were all pupils valued? All teachers from both countries demonstrated genuine warmth and empathy towards and showed respect for the pupils in their classrooms in all respects.

ISTOF 5.2 – Did the teacher initiate active interaction and participation? Chinese teachers were more able to engage every pupil in productive work which generally took pupils less time and was well completed, whereas most English teachers failed to do so and had to re-teach individuals and groups of pupils for multiple times what had been taught because pupils consistently encountered difficulties during independent work. Moreover, across classrooms, Chinese teachers asked more questions than their English colleagues, and Chinese pupils were given more opportunities to answer questions than their peers in England. Chinese teachers' questions tended to follow a coherent logical flow – one led to another, which nurtured deeper thinking amongst pupils; English teachers' questions tended to be dotted with little or loose connections and pose little challenge to pupils.

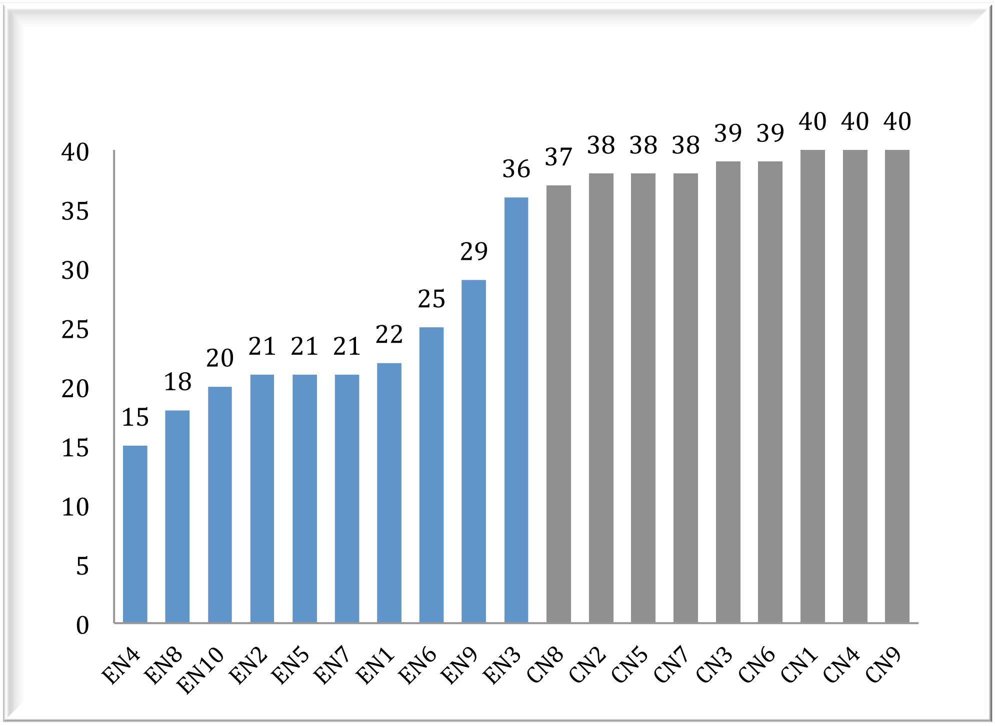


Figure 4.13 ISTOF 5 – Classroom climate

ISTOF 5.3 – Did the teacher interact with all pupils? All teachers showed their effort in involving all pupils particularly those who did not voluntarily participate in classroom activities. Nevertheless, English teachers applied more time interacting with individuals, which hence limited their degree of interacting with *all* pupils in a lesson. The Chinese approach, which featured more proportion of time on whole-class interaction, on the other hand, maximised frequencies of teacher-pupil and pupil-pupil interaction.

ISTOF 5.4 – Did the teacher communicate high expectations? Chinese teachers were more likely to pose reasonable challenge – higher than pupils' actual level but achievable after a little effort – to pupils and praise them for realising their potential. Though English teachers were generous in praising pupils, they often asked easy and less challenging questions and *fed* pupils with too many hints and tips while waiting for answers, thus lowering the degree of challenge they could have posed to them. Thus, the expectation from Chinese teachers to their pupils tended to be much higher than the case of England. Chinese pupils were more motivated, therefore being more confident to meet bigger challenges than their peers in England.

To summarise, the focus of the *classroom climate* domain is more about whether the teacher seeks to build the atmosphere where active learning is nurtured and pupils are motivated to attain more advanced knowledge and seek deeper mathematical thinking. In this respect, Chinese teachers tended to do better than English teachers.

Next section is on the last ISTOF domain, *classroom management*, which carries 3 indicators and 7 behavioural scales.

4.2.2.6 **ISTOF 6 – Classroom management**

In the domain of classroom management, there are seven 5-point scales under the umbrellas of three indicators. Teachers in England scored much less than those in China. The former got a mean of 18.1 (SD=6.4), whereas the latter had an average of 34.2 (SD = 1.7) (Figure 4.14). The difference of scoring between countries also partially explained the better discipline in Chinese classrooms than in English classrooms. There was a statistically significant and strong relationship between nationality and teacher scores in this dimension (t (17) = 7.3, p < .001, d = 3.98 > 1). Detailed accounts are given in the following paragraphs with regard to each indicator.

ISTOF 6.1 – Was learning time maximised? Teachers all started their lessons on time. English teachers however were unable to ensure that all pupils were fully involved in learning activities until the end of the lesson, mainly because they had to stay at the carpet to provide possible re-inputs during the independent-work time. Chinese teachers were conscious of the state of both individuals and the whole class and were able to keep teaching and learning processes more engaging and pupils on task. English teachers did not take many actions to minimize disruption – they seemed quite tolerant with a certain degree of chaos particularly during the periods for independent work. Neither did Chinese teachers take many of such actions overtly, because they did not have to. They had arranged the teaching and learning process in a way that disruption had been covertly prevented.

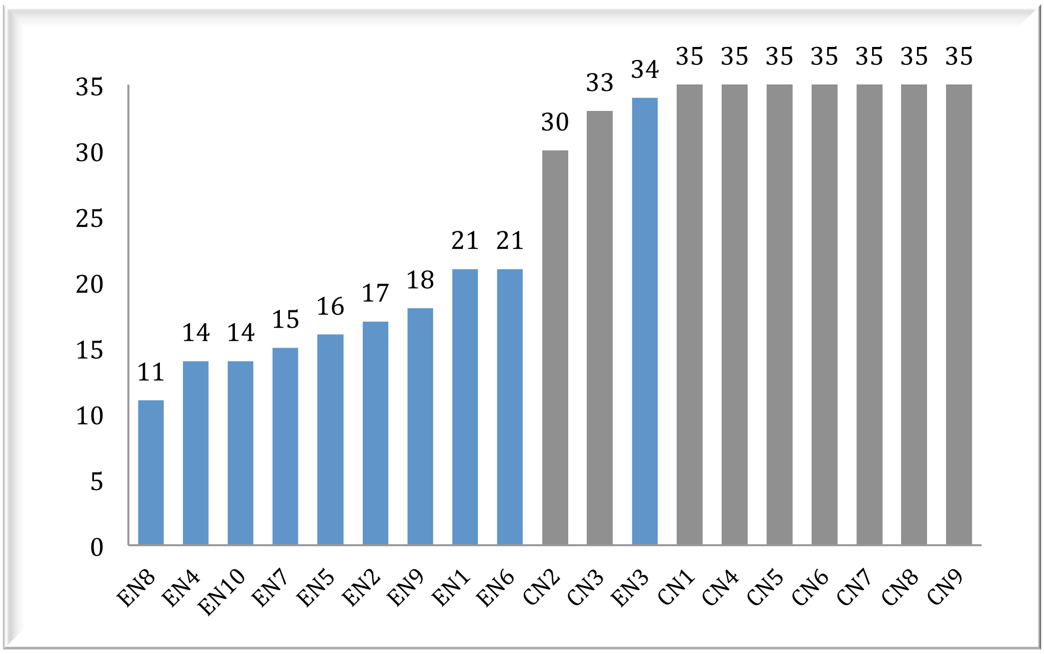


Figure 4.14 ISTOF 6 – Classroom management

ISTOF 6.2 – Were clear rules evident? This dimension looks at two aspects (i.e. 2 scales) of clear rules, clarity of getting help from the teacher and clarity of pupils' options after finishing assignments. Firstly, in the English classrooms, the most efficient way of getting help was to go to the carpet and find the teacher; sometimes pupils had to queue up; a few of them sometimes raised their hands to call for help, but it seemed to take even longer time for the teacher to spot their hands in the air. In the Chinese classrooms, teachers seemed very alerted both visually and mentally and ready to capture any moves in the class immediately; raising hands was therefore the most popular and quickest way for pupils to get attention and help from the teacher.

Secondly, in the English classes, there was a lack of clarity about what options were available when the pupils finish their assignments; there was however clarity about what options were available when the pupils encountered difficulties – to approach the teacher on the carpet. In the Chinese classes, teachers always announced clearly what options were available to those who completed tasks earlier than others; sometimes the option was to check their answers with desk-mates; sometimes it was to sit upright and check their own answers so the teacher could know they were done simply by looking at their sitting postures; brief praise would be given to them if they completed earlier and still managed to behave well while checking their work.

ISTOF 6.3 – Were misbehaviours and disruptions effectively dealt with? After the short whole-class teaching time, English teachers were always busy interacting with individuals or a group of individuals on the carpet, so a small amount of non-academic behaviours among pupils were often not spotted and dealt with. There were quite few misbehaviours and disruptions in the Chinese classrooms; whenever they appeared, Chinese teachers dealt with them quickly and smoothly. All classes from both countries seemed to have their own conventions and rules regarding intolerant misbehaviours and disruptions.

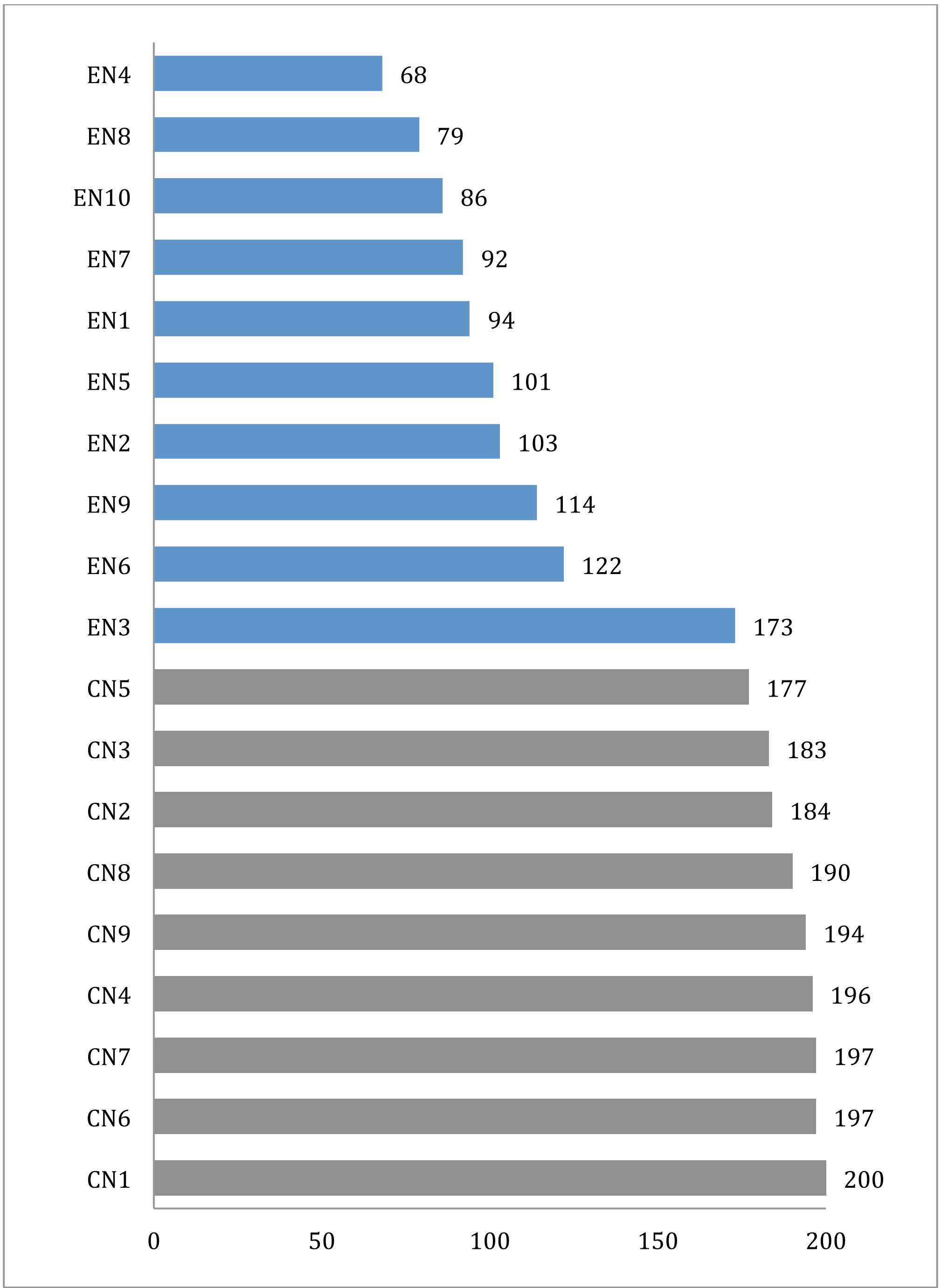


Figure 4.15 ISTOF – Overall scores across classrooms

Having looked into details of the ISTOF results, we will now step back to see the whole picture that the ISTOF measures drew about the quality of mathematics teaching in two countries.

4.2.2.7 Summary on the ISTOF measures

Overall, for the whole instrument, English teachers scored less than Chinese teachers (Figure 4.15).

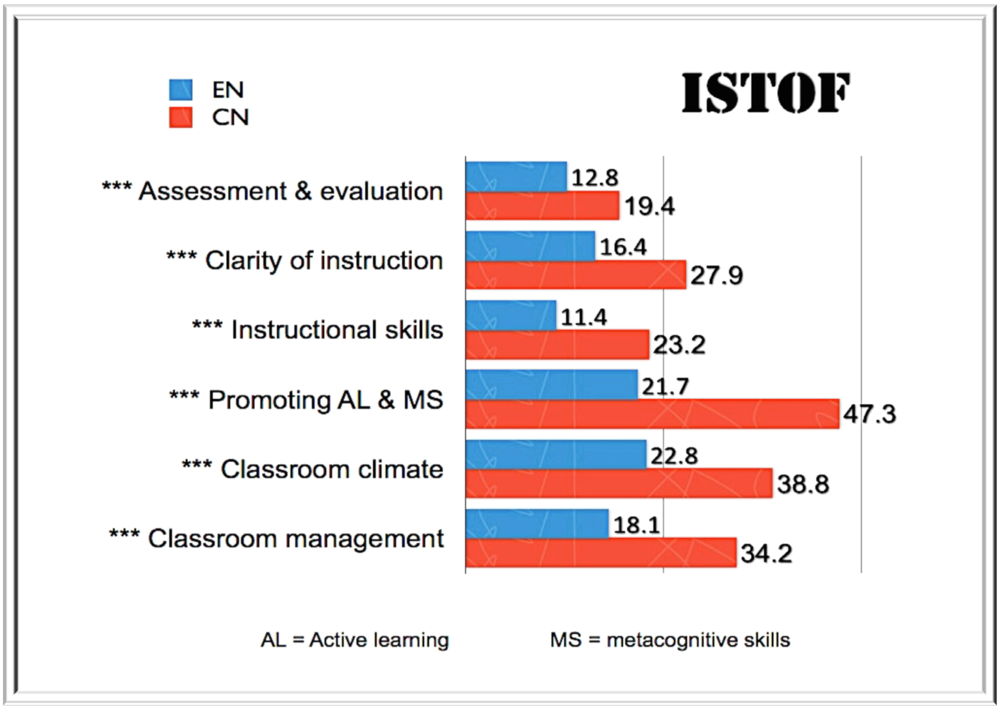


Figure 4.16 The international comparison of six ISTOF scores

English teachers achieved an average of 103.2 (*SD* = 29.2), whilst Chinese teachers had a mean of 190.9 (*SD* = 7.9) (Figure 4.15). The relationship between nationality and the sum of teacher ISTOF scores was statistically significant and strong (t (17) = 8.7, p < .001, d = 4.7 > 1). As shown in Figure 4.16, at the international level, English teachers on average achieved lower scores than did their Chinese colleagues on ISTOF scales.

Now that we have looked into both detailed results and the holistic pattern of the ISTOF evaluation, next, the results of teacher questionnaire will be presented to generate background information about the key player of mathematics teaching.

4.3 Teacher questionnaire (M2.1)

This section mainly presents results of five domains of the teacher questionnaire: (1) *teacher background*, (2) *professional development*, (3) *teacher appraisal and feedback*, (4) *teaching practice, beliefs and attitudes* and (5) *teaching in the observed class*.

4.3.1 Teacher background

As follows is background information about teachers' *demographic information*, such as gender, age and level of education, *work related information*, *hours spent in a typical school week*, and *length of teaching experience*.

4.3.1.1 Demographic information

Teacher gender. In the teacher sample from both countries, there were more females than males. The number of male teachers was three in each country. In both countries, the number of female teachers approximately doubled that of male teachers (Table 4.6) Unsurprisingly, there is no inter-country difference regarding the proportions of teacher genders ($X^2(1, N = 19) = .02, p = .88$).

Teacher age. Eight out of nine Chinese teachers and nine out of ten English teachers were between the ages of 25 and 39. There was only one teacher in England being under 25 and one teacher in China aged 40 to 49. Countries were not significantly different from each other in terms of their teachers' ages ($X^2(3, N = 19) = 2.1, p = .55$).

Table 4.6 The distribution of teacher genders

Nation	*N of Females (percentage)*	*N of Males (percentage)*
England	7 (70%)	3 (30%)
China	6 (66.7%)	3 (33.3%)

Highest level of formal education. Except for one English teacher whose highest degree was Masters, all teachers had a Bachelor's degree in both countries. The nationality of teachers did not have any connection with the highest levels of education they had ($X^2(1, N = 19) = 1.257, p = .262$).

Summary. There were no statistically significant differences found in every surveyed aspect of teachers' demographic information between countries. Next comes teachers' work related information.

4.3.1.2 Work related information

Teacher employment status – school. All the nineteen teachers were working full time at the schools where data were collected, and none of them were also working at another school simultaneously. The nationality did not make any significant difference upon teachers' full-time status at the participating schools. Nor did it impact on their possibility of working for two schools at the same time.

Teacher employment status – contract length. Three Chinese teachers were permanently employed, five had a fixed term contract for more than a year, and one was on the contract of less than a year. Nine English teachers were permanently employed, and one had a less-than-a-year contract. The contact lengths of teachers were significantly and strongly influenced by their nationalities ($X^2(2, N = 19) = 7.97, p = .019 < .05, \varphi = .65$).

Summary. All teachers from both countries were working full time in the participating schools. Nonetheless, there was a significant difference of teacher contract length between two countries, in that most English teachers were permanently employed and that most Chinese teachers were on a fixed-term contract.

4.3.1.3 Hours spent in a typical school week

Teachers' time spent on actual teaching and lesson preparations and administrative and/or other duties was surveyed.

Hours for teaching and lesson preparations. Through the t-test for independent samples, a significant difference was found between Chinese teachers and English teachers regarding the amount of time spent in *teaching* (#8a) and *planning lessons /marking pupil work* (#8b) (Table 4.7).

Table 4.7 Teacher schedules in a typical school week (by hour)

Item #	*Nation*	*N*	*Mean*	*SD*	*t*	*df*	*p*	*d*
8a	CN	9	10.44	4.13	9.5	20	.000	4.1
***ΔΔ	EN	10	25.90	3.48				
8b	CN	9	12.89	3.41	5.6	20	.000	2.3
***ΔΔ	EN	10	21.00	3.56				
8c	CN	9	4.89	.33	-1.735	12.454	.107	-
	EN	10	3.30	2.16				
8d	CN	3	3.00	1.00	-.766	5	.478	-
	EN	3	3.00	1.73				

Notes:
*** = Significant at the .001 level ΔΔ = Strong relationship

Both Chinese teachers' average teaching hours and planning hours approximately halved those for English teachers. Nonetheless, the differences between teaching hours and planning/marking hours suggest that Chinese teachers on average spent slightly more proportions of time in planning lessons and marking pupils' work than did English teachers. As will be presented in Section 4.3.4.5, English teachers all teach multiple subjects, whereas Chinese teachers almost only teach mathematics. This meant that English teachers had to split their time across different subjects. It is thus not clear how much time English teachers might actually spend teaching and planning/marking in the subject of mathematics.

Hours for administrative and/or other duties. There was not much difference in teachers' time spent on *administration* (#8c) or *other commitments* (#8d) between countries (also see Table 4.7). Three English and two Chinese teachers reported that they also had other duties: Teachers EN4 and EN5 both had a role in after-school clubs; Teacher EN6 was a year leader so she had management duties; Teacher CN7 took the responsibility of managing the class during lunchtime and "Sunshine PE" club after school three times a week at her school; Teacher CN9 was involved in the "Happy Wednesday" club and the Guzheng[3] club.

Summary. Overall, English teachers spent considerably more time in both *teaching* and *planning lessons/marking pupil work* than did Chinese teachers. The results suggested that the working load for English teachers was heavier than for Chinese teachers and that Chinese teachers allocated slightly more proportion of time on planning/marking homework than did English teachers. It is worth investigating in future the interconnection between teachers' planning/marking time and teaching and learning outcomes in mathematics.

---

[3] 古筝：an ancient Chinese musical instrument with 21 to 25 strings

4.3.1.4 Length of teaching experience

As follows, the length of teaching experience is informed by two aspects: (1) the total length of experience as a teacher and (2) years into the current position.

Total length of experience as a teacher. As shown in Table 4.8, English teachers' experience ranged from one to 15 years; Chinese teachers' experience had a slightly wider distribution – one to 20+ years. Nonetheless, the difference between two countries was not statistically significant ($X^2(5, N = 19) = 4.63$, $p = .46 > .05$).

Table 4.8 A comparison of the length of teaching experience

	First year	*1-2 years*	*3-5 years*	*6-10 years*	*11-15 years*	*16-20 years*	*20+ years*	*Total*
CN	1	0	2	2	1	2	1	9
EN	3	0	4	2	1	0	0	10
Total	4	0	6	4	2	2	1	19

Table 4.9 The length of teaching experience in current schools

	First year	*1-2 years*	*3-5 years*	*6-10 years*	*11-15 years*	*16-20 years*	*20+ years*	*Total*
CN	2	0	5	1	0	0	1	9
EN	4	0	3	2	1	0	0	10
Total	6	0	8	3	1	0	1	19

Years into the current position. Table 4.9 suggests that most teachers from both countries were either in their first year or 3rd-5th year teaching in the schools where their lessons were observed for the study. There is no international difference in the length of teachers' teaching experience at the participating schools ($X^2(4, N = 19) = 3.46, p = .48 > .05$).

Summary. The length of teachers' experience had no significant difference across countries, and the majority of the teachers in the sample were in their 3rd to 5th years of teaching in current schools. Though Chinese teachers tended to fall evenly into the time spans, this did not make a significant difference. Judging by the differences between Table 4.8 and Table 4.9, English teachers were more likely to stay in the same school where they started their career, while less likely were Chinese teachers. Three Chinese teachers with the experience of 16-20 or 11-15 years teaching had moved to the current school for 3 to 5 years, whereas only one English teacher with a total of 3-5 years' experience had recently moved to the current school.

With teacher background information in mind, we will then go on to look at details about teacher professional development.

4.3.2 Professional development

In this domain, four groups of questions were given to teachers regarding: (1) *professional development during the past 18 months and its impact*, (2) *time and money for professional development*, (3) *less formal professional development* and (4) *teacher professional development needs*. Results are presented sequentially as follows.

4.3.2.1 Professional development during the past 18 months and its impact

The following paragraphs provide information about teachers' frequencies and perceived impacts of taking part in professional development activities, such as *courses/workshops*, *education conferences/seminars*, *qualification programmes*, *observation visits to other schools*, *teacher networks*, *research activities*, and *monitoring and/or peer observation and coaching*.

Courses/workshops. All nineteen teachers have participated in courses/workshops during the past 18 months. Hence, there is no apparent difference between countries. All nine Chinese teachers thought the courses/workshops they took part in had a moderate impact on their development as a teacher. Seven English teachers (70%) perceived a moderate impact, and three others (30%) a large impact. Although the ratings of the impact varied slightly across countries, there is in fact no significant relationship between nationality and the corresponding level of perceived impact ($X^2(1, N = 19) = 3.21, p = .07 > .05$).

Education conferences/seminars. Nine (90%) out of ten English teachers have not participated in education conferences/seminars during the past 18 months, whereas eight (88.9%) out of nine Chinese teachers have done so. It is quite obvious that teachers' possibility of attending education conferences/seminars varied considerably between countries ($X^2(1, N = 19) = 11.83, p = .001$). Moreover, there is also a strong effect size ($\varphi = .79$). The only English teacher, who had participated in conferences/seminars, rated the activity as a small impact. Of the eight Chinese teachers who had also made their way to conferences/seminars, two perceived a small impact, four a moderate impact, and two a large impact.

Qualification programmes. The majority of teachers in each country – nine (90%) out of ten in England and six (66.7%) out of eight (one missing case) in China – have not participated in qualification programmes during the last 18 months. There is no significant difference between countries in the participation rate ($X^2(1, N = 18) = .72, p = .4 > .05$). In terms of impact, all three teachers – one English and two Chinese – perceived that their participation in qualification programmes had promoted their professional development as teachers, with one English and one Chinese rating this as a small impact and the other Chinese a large impact. Unsurprisingly, there is no significant difference between countries regarding the perceived impact of taking part in qualification programmes ($X^2(1, N = 18) = .75, p = .39 > .05$).

Observation visits to other schools. Almost all teachers (nine out of ten English and all nine Chinese) have taken observation visits to other schools during the last 18 months. This again does not show any significant relationship between nationalities and the possibility of taking part ($X^2(1, N = 19) = .95, p = .33 > .05$). Regarding such observation visits, four English teachers and one Chinese teacher perceived a small impact, three English and six Chinese teachers a moderate impact, and two teachers from each country a large impact. Even though Chinese teachers tended to perceive higher levels of impact, the difference of perceived impact between countries is insignificant ($X^2(2, N = 19) = 2.8, p = .25 > .05$).

Teacher networks. Over the last 18 months, four out of ten English teachers (40%) had taken part in a network of teachers for the purpose of professional development, and so had eight out of nine Chinese teachers (75%). The international difference was statistically significant with a moderate to strong effect ($X^2(1, N = 19) = 4.87, p = .03 < .05, \varphi = .51$). The perceived impact of participating in teacher networks was however not significantly different between countries ($X^2(2, N = 19) = 2.25, p = .33 > .05$). Of four English teachers who participated, one received a small impact, and three a moderate impact. Of eight Chinese teachers who did so, two rated it as a small impact, three a moderate impact, and three a large impact.

Research activities. When being asked about their involvement into individual/collaborative research activities, 44.4% of Chinese teachers (4 out of 9) answered 'Yes', whereas all English teachers chose 'No'. There thus appears to be a significant difference between countries ($X^2(1, N = 19) = 5.63, p = .02 < .05$) and a moderate to strong effect size ($\varphi = .54$). It is a widely known fact in China that teachers' research-related activities and publications are an important appraisal criterion. Of the four Chinese teachers who have conducted research, three rated such activities as a moderate impact upon their development as a teacher, and one a large impact.

Mentoring and/or peer observation and coaching. Eight out of ten English teachers and all nine teachers have conducted mentoring and/or peer observation and coaching during the last 18 months. Statistically, there is no apparent difference between countries ($X^2(1, N = 19) = 2.01, p = .16$). All seventeen teachers, who have participated in this type of activities, received a positive impact. Four English teachers rated it as a moderate impact, and so did five Chinese teachers. Four teachers from each country perceived a large impact. No significant relationship between nationalities and impact ratings was found ($X^2(1, N = 19) = .052, p = .82 > .05$).

Summary. A significant difference was found between countries in three of the above seven types of professional development activities. Chinese teachers participated more frequently than did English teachers in *education conferences/seminars*, *teacher networks* and *research activities*.

4.3.2.2 Time and money for professional development

Time for professional development. All teachers except a missing case from China reported the number of days they had spent on professional development during the last 18 months. English teachers had on average spent 5.7 days (*SD* = 3.7), while Chinese 44.4 days (*SD* = 23.5) (Figure 4.17). The t-test for independent samples indicated that the difference of time for professional development between countries was significant at the level of .01 ($t(7) = 4.6$, $p = .002 < .01$), and the calculation of Cohen's d further indicated that the relationship between nationality and days of professional development was strong ($d = 2.38 > 1$). In China, while major professional development programmes often run during the summer holiday (two months) or winter holiday (about three weeks), many other virtual programmes are also available during term days. It is very common and trendy across Chinese schools, particularly those in cities, that teachers and school leaders are joining various training courses each year. It is not clear whether those programmes/courses are helpful, but it is clear that teachers, school leaders and even the whole society are eager for new ideas.

Time for compulsory professional development. Of time spent on these professional development, an average of 4.2 days (*SD* = 3.2) was compulsory for English teachers, and 34.5 days for Chinese teachers (*SD* = 18.3).

Professional development during work hours. All English teachers had received scheduled time for undertaking professional development that took place during regular work time, and so did seven out of nine Chinese teachers. There were one Chinese teacher who did not have scheduled time for this and another who took professional development that did not take place during regular work hours. No significant relationship was found between teachers' nationalities and their possibilities of receiving this type of professional development ($X^2(2, N = 19) = 2.3$, $p = .33 > .05$).

Paid or not? Of all teachers from both countries (one English case missing for this question), only one Chinese teacher reported receiving a supplement salary. Four English teachers and eight Chinese teachers took the professional development taking place outside regular work hours without a salary supplement for it. Five English teachers did not take part in professional development outside regular work hours. The international difference is significant ($X^2(2, N = 18) = 7.3$, $p = .03 < .05$), and the effect is strong ($\varphi = .64$).

Summary. As shown in Figure 4.17, Chinese teachers spent much more days on professional development than did their English colleagues during the past 18 months. This indicated that Chinese teachers had more opportunities to grow professionally, which may indirectly contribute to the improvement of their classroom practices.

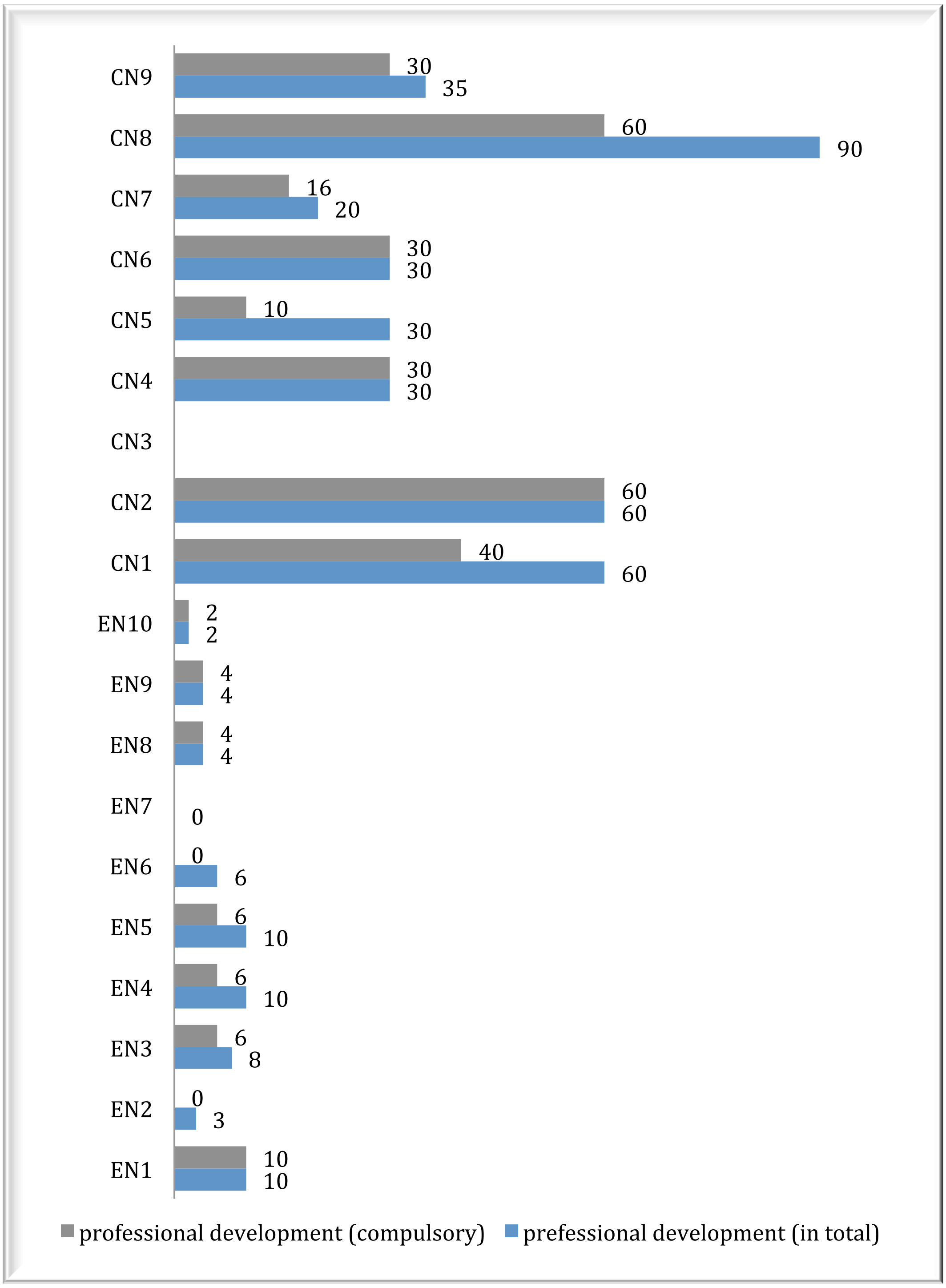


Figure 4.17 Days of professional development undertaken

The following section provides information as to whether or not teachers had carried out less formal professional development, such as *reading professional literature* and *informal dialogue with colleagues on how to improve teaching*.

4.3.2.3 Less formal professional development

Reading professional literature. When being asked whether they had read professional literature during the last 18 months, 3 English teachers (30%) said yes, and so did 8 Chinese teachers (88.9%). All the remaining teachers (7 from England and 1 from China) had not done so. Thus, there is a significant difference between two countries in this aspect ($X^2(1, N = 19) = 6.7, p = .009 < .01$) with a strong effect ($\varphi = .6$). Of the 11 teachers who had read professional literature during the past 18 months, all three English teachers and only one Chinese teacher indicated that this type of activity had a small impact. Four Chinese teachers perceived a moderate impact, and three a large impact. There is a significant difference between two countries regarding the perceived impact of reading professional literature ($X^2(2, N = 19) = 7.2, p = .03 < .05$), with a very strong effect ($\varphi = .81$). Chinese teachers were more likely to perceive a larger impact from this than did English teachers.

Informal dialogue with colleagues on how to improve teaching. Almost all teachers (18 out of 19) had engaged in informal dialogue with their colleagues on how to improve their teaching during the past 18 months. Only one English teacher had not done so. It seems obvious that there is no significant difference between countries on this type of activity ($X^2(1, N = 19) = .95, p = .33 > .05$). Among nine English teachers who had engaged in such dialogue, four received a moderate impact upon their development as a teacher and five a large impact; among nine Chinese teachers who had also engaged in such dialogue, five received a moderate impact, and four a large impact. It is almost observable that there is no international difference in teachers' perceived impact of this type of peer communication on their development as teachers ($X^2(1, N = 19) = .22, p = .64 > .05$).

Summary. The results indicated that most Chinese teachers had read professional literature during the past 18 months, whereas over two thirds of English teachers had not. Chinese teachers tended to feel a higher level of benefit in doing so, whereas English teachers tended to perceive a lower level of such benefit.

4.3.2.4 **Teacher professional development needs**

In this question, teachers were asked to rate their needs in eleven areas of professional development. Ratings were made on a four-point scale providing the following choices: (1) no needs at all, (2) low level of needs, (3) moderate level of needs and (4) high level of needs. Teachers' expected professional development and obstacles that prevented them from reaching the expected level in the past 18 months were also surveyed.

Needs in contents and performance standards in teachers' main subject field. Of all nineteen teachers, eighteen teachers expressed various levels of needs, with just one English teacher demanding no need at all. Five English teachers had a low level of such needs, and four others had a moderate level of needs. The nine Chinese teachers' needs were evenly distributed into the three levels – low, moderate, and high. There is no significant difference found between countries regarding teachers' such needs ($X^2(3, N = 19) = 4.6, p = .2 > .05$).

Needs in pupil assessment practices. Teachers all expressed their needs at different levels. Three English and two Chinese teachers had a low level of needs; seven English and six Chinese teachers had a moderate level of needs; a Chinese teacher had a high level of such needs. The most rated level was the moderate level for both countries. Unsurprisingly, there is no significant relationship between teachers' nationalities and their needs in developing assessment practices ($X^2(2, N = 19) = 1.2, p = .54 > .05$).

Needs in classroom management. Four English teachers had no needs at all, five had a low level of needs, and one had a moderate level of needs. One Chinese teacher had a low level of needs, seven a moderate level of needs, and one a high level of needs. There was a significant and very strong relationship between teachers' nationalities and their levels of needs in this aspect ($X^2(3, N = 19) = 12.2$, $p = .007 < .01, \varphi = .8$).

Needs in knowledge and understanding of teachers' main subject field. English teachers tended to choose lower demand, with one expressing no needs at all, eight having a low level of needs and one a moderate level; Chinese teachers had higher demands: one demanding a low level of needs, three moderate and five high. There is a significant and very strong relationship between nationalities and teacher development needs in subject knowledge ($X^2(3, N = 19) = 12.4$, $p = .006 < .01$, $\varphi = .81$). It is interesting to see that Chinese teachers expressed more development needs in the subject matter than English teachers, as the literature indicates Chinese mathematics teachers demonstrate profounder subject matter knowledge than their Western counterparts (for example than American teachers, see Ma, 1999).

Needs in developing pedagogical content knowledge. Nine English teachers expressed a low level of needs, and one moderate; one Chinese teacher had a low level of needs, two moderate, and six high. Again, like development needs in the area of subject knowledge, Chinese teachers expressed higher levels of needs in instructional practices than did their English colleagues. There is a significant and very strong relationship between teachers' nationalities and their professional development needs in this area ($X^2(2, N = 19) = 12.7$, $p = .002 < .01$, $\varphi = .82$).

Needs in ICT skills for teaching. Only one English teacher had no needs at all, three had a low level of needs, five moderate, and one high; three Chinese teachers had a low level of needs, four moderate, and two high. There was no significant cross-national difference in teachers' needs in ICT skills ($X^2(3, N = 19) = 1.4$, $p = .71 > .05$).

Needs in teaching pupils with special educational needs (SEN). All teachers had expressed their needs at certain levels: three English and two Chinese teachers had a low level of needs; five English and four Chinese had a moderate level of needs; two English and three Chinese demanded a high level of needs. There was no statistically significant difference between countries in teachers' development needs in teaching children with SEN ($X^2(2, N = 19) = .46$, $p = .8 > .05$).

Needs in pupil discipline and behaviour problems. Three English teachers thought they did not need any professional development in this aspect, and seven had a low level of needs. On the contrary, only one Chinese demanded a low level of needs, four moderate, and four high. The difference between countries was statistically significant ($X^2(3, N = 19) = 15.5$, $p = .001 < .01$) with a very strong effect ($\varphi = .9 > .8$). Chinese teachers tended to have higher needs.

Needs in school management and administration. In this aspect, the distribution in England was quite similar to that in China: one English and two Chinese teachers had no needs at all, four English and four Chinese teachers had a low level of needs, and four English and two Chinese expressed a moderate level of needs; one in each country expected a high level of such needs. Statistically, there was no significant difference between countries ($X^2(3, N = 19) = .95$, $p = .8 > .05$).

Needs in teaching in multicultural settings. One English teacher demanded no needs at all, four a low level of needs, and five moderate. One Chinese teacher had no need at all, two low, three moderate, and three high. There was no significant cross-national difference in teachers' such needs ($X^2(3, N = 19) = 4.1$, $p = .25 > .05$).

Needs in pupil counselling. Chinese tended to express higher levels of needs, with six expressing a moderate level of needs and three high. One English teacher had no needs at all, five low, and four moderate. There was a significant difference between two countries in teachers' development needs in pupil counselling ($X^2(3, N = 19) = .94$, $p = .025 < .05$) with a strong effect ($\varphi = .702$).

More participation? To summarise, teachers' willingness to participate in more professional development than they did during the past 18 months were half and half for the case of England. Three Chinese teachers wanted more than they did, and five did not want more, with one missing case. There was no significant international difference for this question ($X^2(1, N = 18) = .28$, $p = .596 > .05$).

Obstacles to more participation. For those who wanted more professional development than they did, two Chinese teachers attributed the obstacle to "I did not have the pre-requisites (e.g. qualifications, experiences, seniority)", and one attributed it to "professional development conflicted with my work schedule".

One English teacher referred the reason to "professional development was too expensive / I could not afford it", two thought, "professional development conflicted with my work schedule", one indicated "there was no suitable professional development offered", and one attributed it to a specific reason - workload. Again, there was no significant relationship between nationality and reasons for insufficient professional development ($X^2(4, N = 8) = 5.2$, $p = .272 > .05$).

Summary. Teachers from both countries had no significant differences of needs in the surveyed aspects of professional development, except for *classroom management*, *knowledge and understanding of teachers' main subject field*, *developing pedagogical content knowledge*, *pupil discipline and behaviour problems* and *pupil counselling*. In these five areas, Chinese teachers expected significantly higher needs than English teachers.

Now that we have looked at the background and professional development of teachers in both countries, next section will describe the status of teacher appraisal and feedback.

4.3.3 Teacher appraisal and feedback

This domain focuses on (1) frequency of receiving appraisal and/or feedback, (2) aspects considered for the appraisal and/or feedback, (3) the influence of appraisals on the external circumstances of teachers, (4) the influence of appraisals on the professional self and (5) overall influence of appraisals upon teachers.

4.3.3.1 **Frequency of receiving appraisal and/or feedback**

From the head. The frequencies on which teachers received appraisal and/or feedback from their headteachers seemed more often for English teachers than for the Chinese. Eight English teachers received such feedback three or more times per year, one received it monthly, and another more than once per month. Three Chinese teachers received feedback from headteachers once per year, and six three or more times per year. Despite the difference revealed above, it was not statistically significant ($X^2(3, N = 19) = 5.2, p = .155 > .05$).

From other colleagues. Two English teachers received such appraisal and/or feedback once per year, one twice per year, two three or more times per year, three monthly, and two more than once per month. One Chinese teacher received this type of appraisal and/or feedback less than once every two years, five received this three or more times per year, and three monthly. The frequency was not statistically significant between countries ($X^2(5, N = 19) = 7.25, p = .2 > .05$).

From external individual or body. Two English teachers never received any appraisal/feedback from an external individual or body, two received less than once every two years, one once every two years, four once per year, and one twice per year. One Chinese teacher received such appraisal/feedback once every two years, three once per year, three twice per year, and two three or more times per year. The frequencies for English teachers tended to spread across the whole range of choices, whereas those for Chinese teachers were skewed to the higher end of the scale and between once every two years and three or more times per year. Such a difference between countries, however, was not statistically significant ($X^2(5, N = 19) = 7.1, p = .2 > .05$).

Summary. There are no significant international differences regarding teachers' frequencies of receiving appraisal and/or feedback from the head teacher, colleagues or external individuals/bodies.

4.3.3.2 Aspects considered for the appraisal and/or feedback

Eighteen aspects were provided for teachers to rate the extent to which they think each aspect was considered when they were appraised.

Table 4.10 shows a detailed international comparison of teachers' responses as a result of cross-tabulation analysis in SPSS. It has been found that four aspects were rated significantly differently between countries with a strong effect. These four aspects are: *pupil test scores* (22a), *other pupil learning outcomes* (22c), *direct appraisal of my classroom teaching* (22g), and *teaching in multicultural settings* (22p). Other aspects that did not show significant differences cross-nationally are: *pupil test scores* (22a), *retention and pass rates of pupils* (22b), *pupil feedback on my teaching* (22d), *feedback from parents* (22e), *how well I work with the head teacher and my colleagues* (22f), *innovative teaching practices* (22h), *relations with pupils* (22i), *professional development I have undertaken* (22j), *classroom management* (22k), *knowledge and understanding of my main subject field(s)* (22*l*), *knowledge and understanding of instructional practices (knowledge mediation) in my main subject field(s)* (22m), *teaching pupils with special learning needs* (22n), *pupil discipline and behaviour* (22o), *extra-curricular activities with pupils (e.g. school plays and performances, sporting activities)* (22q), and *other (please specify)* (22r).

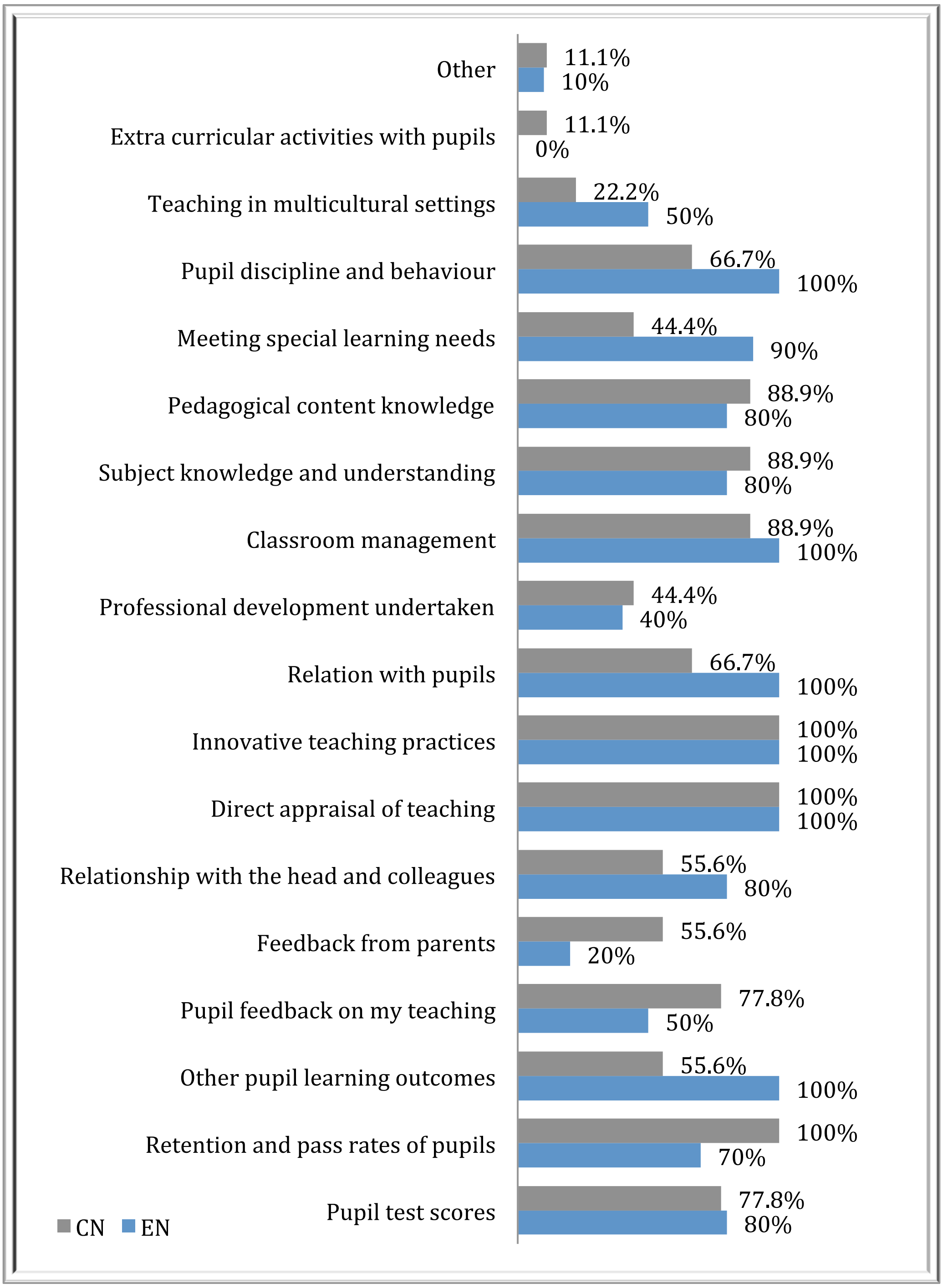
In order to further compare the importance of these aspects as perceived by teachers, the percentages of teachers choosing moderate and high importance were accumulated and shown in Figure 4.18. Two countries seemed to both emphasise two aspects in teacher appraisals: *innovative teaching practices* and *direct appraisal of teaching*. In addition, England also had the highest emphasis on three other aspects: *pupil discipline and behaviour*, *relation with pupils*, and *other pupil learning outcomes*, while China focused most on *retention and pass rates of pupils*, probably due to the crucial role that exam results play in the country.

Table 4.10 Teacher perceived focus of the appraisal/feedback

#	*Nation*	*1*	*2*	*3*	*4*	*5*	X^2	*df*	*p*	*φ*
22a	CN	0	2	0	7	0	10.577	3	.014	.746
*∆	EN	0	0	2	3	5				
22b	CN	0	0	0	6	3	5.463	2	.065	-
	EN	3	0	0	2	5				
22c	CN	0	1	3	5	0	10.086	3	.018	.729
*∆	EN	0	0	0	4	6				
22d	CN	0	2	0	4	3	3.299	4	.509	-
	EN	1	2	2	3	2				
22e	CN	0	2	2	3	2	6.164	4	.187	-
	EN	4	2	2	2	0				
22f	CN	0	1	3	4	1	2.064	3	.559	-
	EN	0	1	1	5	3				
22g	CN	0	0	0	4	5	5.630	1	.018	.544
*∆	EN	0	0	0	0	10				
22h	CN	0	0	0	8	1	.281	1	.596	-
	EN	0	0	0	8	2				
22i	CN	0	2	1	4	2	7.234	3	.065	-
	EN	0	0	0	2	8				
22j	CN	0	3	2	4	0	3.958	3	.266	-
	EN	3	1	2	4	0				
22k	CN	0	0	1	6	2	3.357	2	.187	-
	EN	0	0	0	4	6				
22l	CN	0	0	1	5	3	2.574	2	.276	-
	EN	0	0	2	2	6				
22m	CN	0	0	1	5	3	1.284	2	.526	-
	EN	0	0	2	3	5				
22n	CN	0	1	4	2	2	4.961	3	.175	-
	EN	0	0	1	6	3				
22o	CN	0	0	3	5	1	5.630	2	.060	-
	EN	0	0	0	5	5				
22p	CN	0	2	5	2	0	9.641	4	.047	.712
*∆	EN	3	2	0	4	1				
22q	CN	0	3	4	2	0	7.310	3	.063	-
	EN	5	2	3	0	0				
22r	CN	0	0	0	1	0	2.000	1	.157	-
	EN	0	0	0	0	1				

Notes

1 = I don not know if it was considered
2 = Not considered at all
3 = Considered with low importance
4 = Considered with moderate importance
5 = Considered with high importance
* = Significant at the .05 level
∆ = Strong effect, i.e. fall in the [0.5, 0.8) interval



(moderate to high)

Figure 4.18 Teacher-perceived focus of the appraisal/feedback

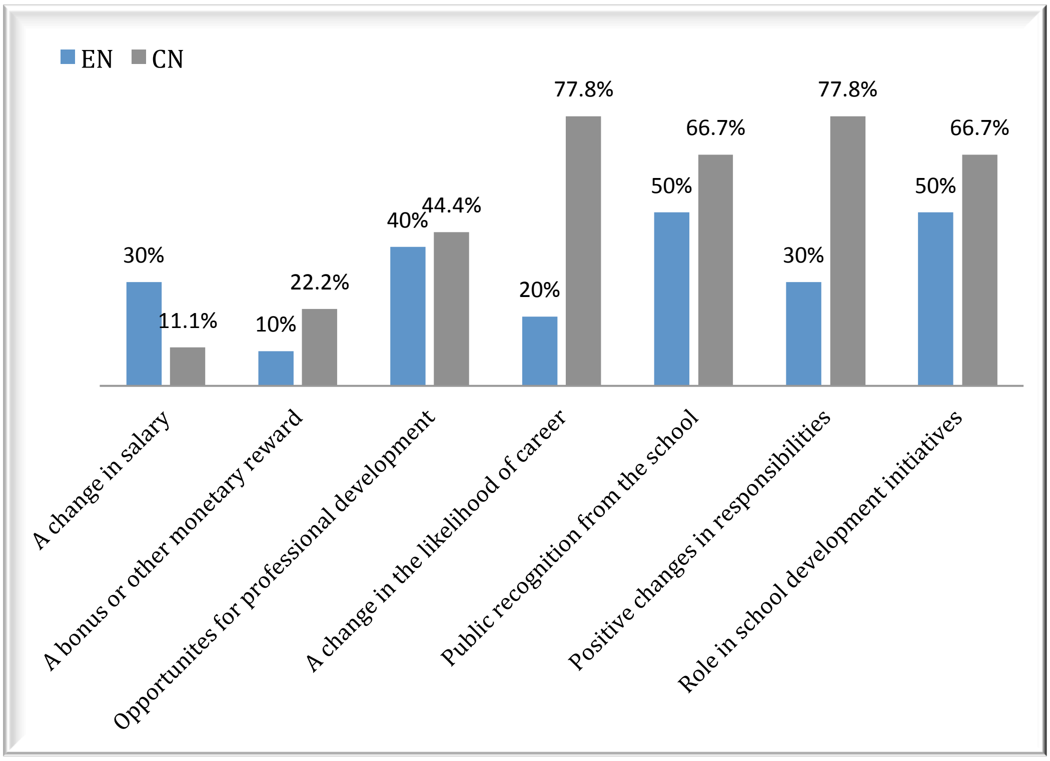
4.3.3.3 The influence of appraisals on the career status of teachers

This section focused on teachers' view on the impact of seven reward- or promotion-related aspects: (1) *a change in salary* (23a), (2) *a financial bonus or another kind of monetary reward* (23b), (3) *opportunities for professional development activities* (23c), (4) *a change in the likelihood of career advancement* (23d), (5) *public recognition from the head teacher and/or your colleagues* (23e), (6) *changes in your work responsibilities that make the job more attractive* (23f) and (7) *role in school development initiatives (e.g. curriculum development group, development of school objectives)* (23g). As shown in Table 4.11, the nationality of a teacher does not make his/her perceived change in any of the seven aspects any different than that of another.

Table 4.11 The influence of appraisals on teachers' career status

#	*Nation*	*1*	*2*	*3*	*4*	X^2	*df*	*p*	*φ*
23a	CN	8	0	1	0	5.630	3	.131	-
	EN	4	3	2	1				
23b	CN	4	3	2	0	5.218	2	.074	-
	EN	9	0	1	0				
23c	CN	3	2	4	0	4.913	3	.178	-
	EN	1	5	2	2				
23d	CN	1	1	4	3	7.836	3	.050	.642
	EN	5	3	0	2				
23e	CN	1	2	4	2	.625	3	.891	-
	EN	2	3	3	2				
23f	CN	0	2	5	2	6.833	3	.077	-
	EN	4	3	1	2				
23g	CN	0	3	6	0	2.497	3	.476	-
	EN	1	4	4	1				

Notes
1 = No change
2 = A small change
3 = A moderate change
4 = A large change



(moderate to large)

Figure 4.19 The impact of appraisals on teachers' career status

The percentage of English teachers that perceived a change in salary after appraisals tripled that of Chinese teachers (Figure 4.19). In each of the remaining six aspects, the proportion of Chinese teachers that rated the appraisal-affected changes as moderate to high was more than that of English teachers. Particularly in three aspects, the proportion of Chinese teachers who perceived a moderate to high change almost doubled that of English teachers. These three aspects are (1) *a bonus or other monetary reward*, (2) *a change in the likelihood of career advancement* and (3) *positive changes in responsibilities*.

4.3.3.4 The influence of appraisals on the professional self

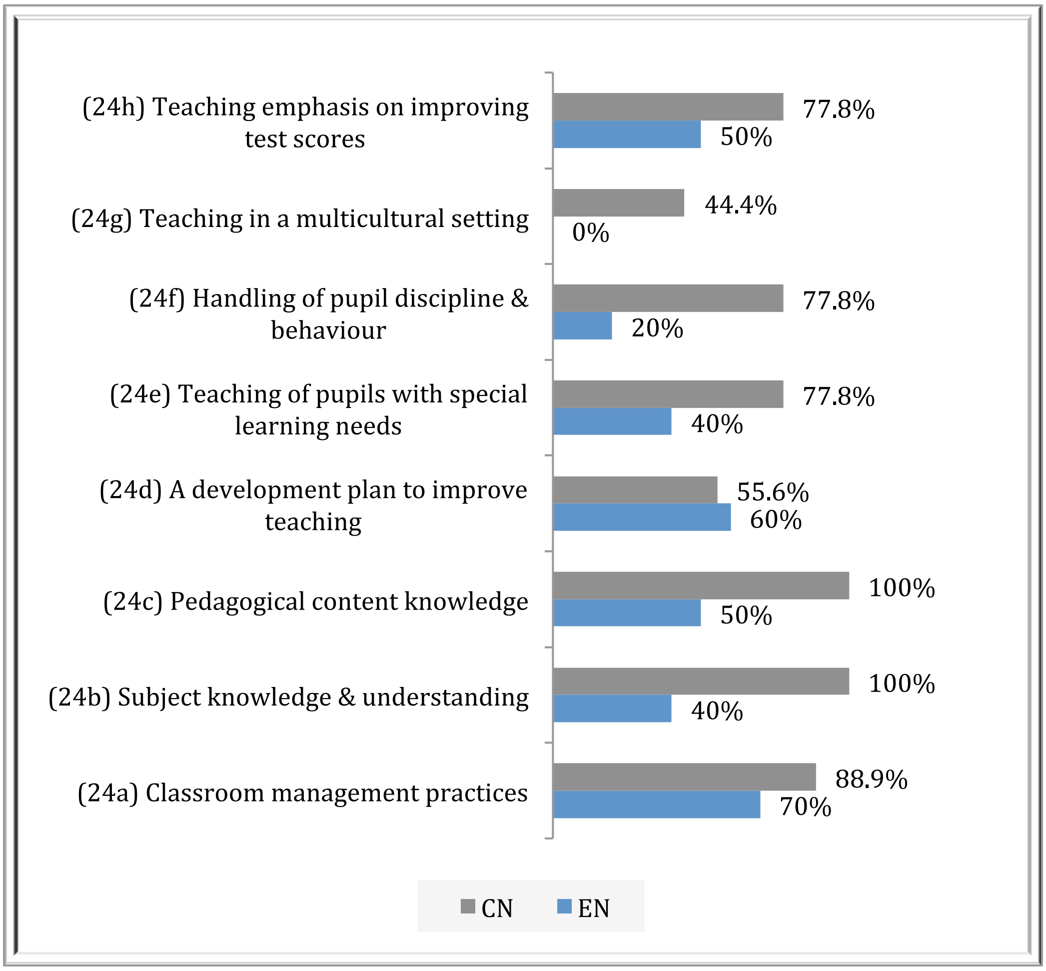
Eight aspects of changes (see Figure 4.20 for details) were asked to measure the extent to which teacher appraisals impacted upon teachers' growth in subject and pedagogical content knowledge and teaching practices.

Table 4.12 The impact of teacher appraisals on the professional self

#	Nation	1	2	3	4	X^2	df	p	φ
24a	CN	0	1	7	1	2.773	2	.250	-
	EN	0	3	4	3				
24b *Δ	CN	0	0	7	2	10.677	3	.014	.750
	EN	2	4	1	3				
24c *Δ	CN	0	0	7	2	7.947	3	.047	.647
	EN	1	4	2	3				
24d	CN	1	3	5	0	2.064	3	.559	-
	EN	1	3	4	2				
24e	CN	1	1	4	3	2.822	3	.420	-
	EN	3	3	2	2				
24f	CN	0	2	5	2	7.253	3	.064	-
	EN	3	5	1	1				
24g *Δ	CN	1	4	3	1	8.614	3	.035	.673
	EN	7	3	0	0				
24h	CN	0	2	6	1	2.554	3	.466	-
	EN	2	3	4	1				

Notes
1 = No change
2 = A small change
3 = A moderate change
4 = A large change
* = Significant at the .05 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval

Table 4.12 indicates that the perceived impact of teacher appraisals varied significantly between countries in the aspects coded 24b, 24c, and 24g, each with a strong effect.



(moderate to large)

Figure 4.20 Impact of teacher appraisals on the professional self

The accumulated percentages of moderate to high ratings (Figure 4.20) show that except for one aspect, *a development or training plan to improve teaching* (24d), more Chinese teachers perceived a positive change than the corresponding proportion of English teachers. In particular, the proportion of Chinese teachers perceiving changes in handling pupil discipline and behaviour problems were approximately four times of that of English teachers. It is also striking to see that

44.4% Chinese teachers perceived a change in teaching in a multicultural setting, whereas none of English teachers has such a perception. Moreover, the Chinese percentage in each of the following three aspects approximately doubled the corresponding English rate: (1) *teaching pupils with special learning needs* (24e), (2) *pedagogical content knowledge* (24c) and (3) *subject knowledge and understanding* (24b). These differences suggest that Chinese teacher appraisal frameworks might have a considerable emphasis on the quality of teaching and teacher knowledge and teachers' influence upon pupils. It might yield insights into the potential indirect impact of teacher evaluation on learning outcomes to carry out cross-national comparisons of teacher appraisal protocols in future.

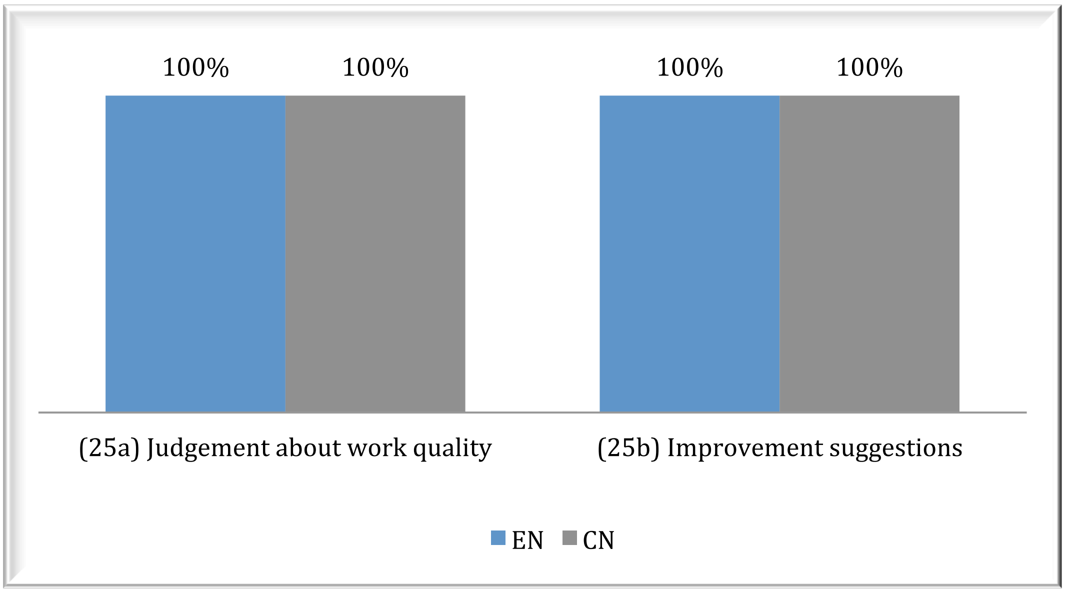
Next, we will look at the holistic view of teachers regarding the appraisal/feedback.

4.3.3.5 Teachers' overall impressions of appraisals

In this section, teachers' overall impressions of appraisals are drawn from four strands of ratings: (1) key messages from the appraisal/feedback, (2) fairness and helpfulness of the appraisal/feedback, (3) direct change brought by the appraisal/feedback and (4) teacher perceived school policy on appraisals.

Key messages from the appraisal/feedback. All teachers from both countries agreed that the appraisal/feedback contained a judgment about the quality of their work (25a) as well as suggestions for improving certain aspects of their work

(25b). Chi-squares for both 25a and 25b were constants, given the 100% rates for both countries (Figure 4.21).



(Yes)

Figure 4.21 Key messages from the appraisal/feedback

Fairness and helpfulness of the appraisal/feedback. All English teachers agreed to a certain degree that the appraisal/feedback they received were *fair* (26a) and *helpful* (26b). One to two Chinese teachers disagreed to a various degree that the appraisal was unfair and/or unhelpful. The international difference regarding the fairness of teacher appraisals was significant at the level of .05 (Table 4.13).

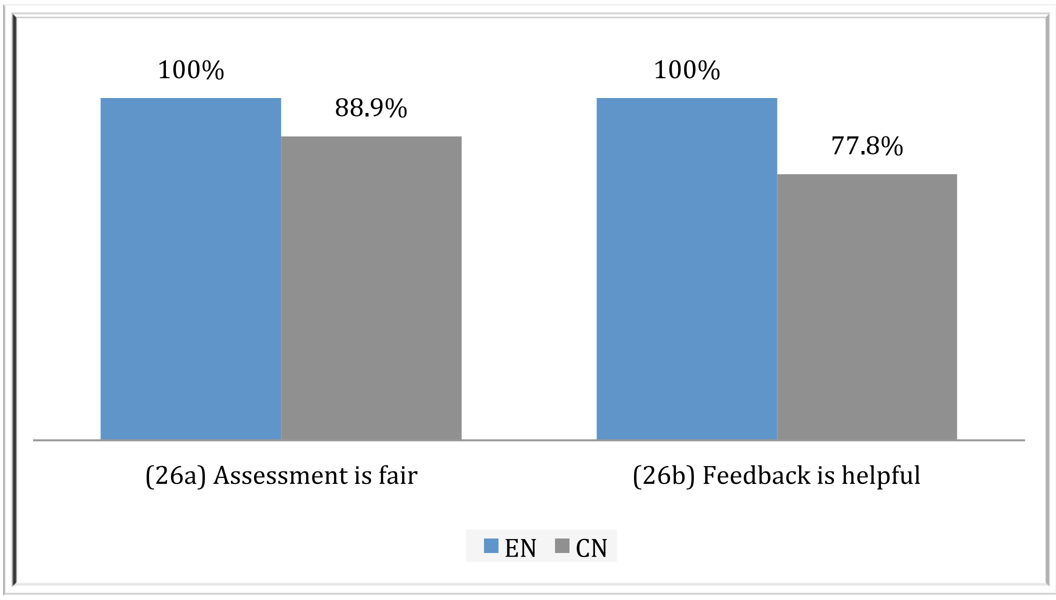
The accumulated percentages in Figure 4.22 further show that, whilst all English teachers had a positive perception of both fairness and helpfulness of the appraisal/feedback, about 1/10 of Chinese teachers did not feel it was fair, and approximately 1/4 of them did not think the appraisal/feedback was helpful.

Table 4.13 Fairness & helpfulness of teacher appraisals

(% Agree & Strongly Agree)

#	*Nation*	*1*	*2*	*3*	*4*	X^2	*df*	*p*	*φ*
26a *Δ	CN	1	0	7	1	9.195	2	.01	.696
	EN	0	0	2	8				
26b	CN	1	1	5	2	4.070	3	.463	-
	EN	0	0	4	6				

Notes
1 = Strongly disagree
2 = Disagree
3 = Agree
4 = Strongly agree
* = Significant at the .05 level
Δ = Strong effect, fall in the [0.5, 0.8) interval



(Agree & Strongly Agree)

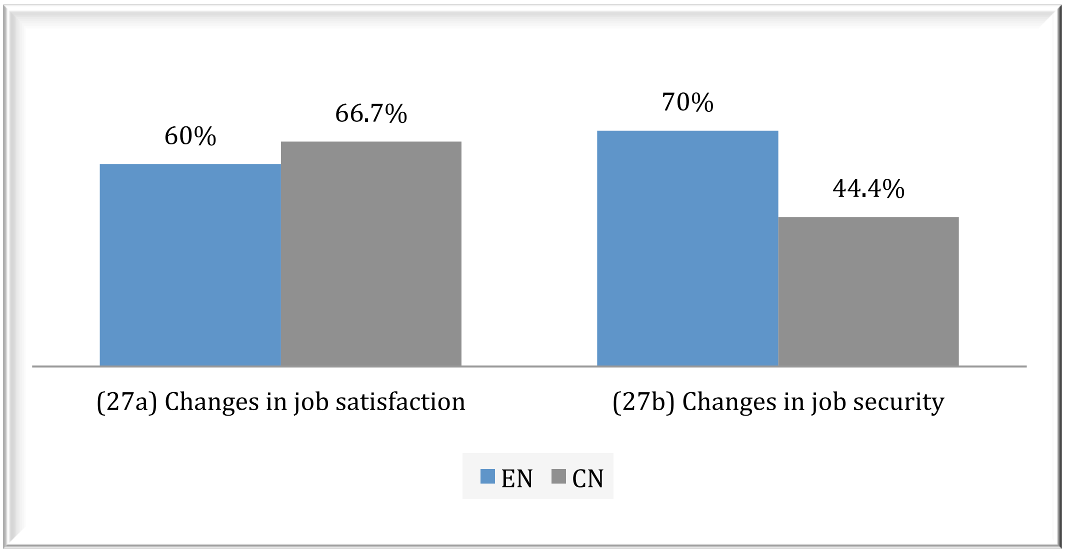
Figure 4.22 Fairness and helpfulness of teacher appraisals

Direct changes brought by the appraisal/feedback. The international difference in perceived changes in both *job satisfaction* (27a) and *job security* (27b) was not significant, as shown in Table 4.14. A varying amount of teachers either perceived a decrease or no change in these two aspects. It suggests a possible distraction that a poor appraisal might pose to teachers at work. No significant difference was found between two countries in these two aspects.

Table 4.14 Changes of job satisfaction & security led by appraisals

#	Nation	1	2	3	4	5	X^2	df	p
27a	CN	0	1	2	4	2	.492	3	.921
	EN	0	1	3	3	3			
27b	CN	0	0	5	3	1	2.254	2	.324
	EN	0	0	3	3	4			

Notes
1 = A large decrease
2 = A small decrease
3 = No change
4 = A small increase
5 = A large increase



(Small & Large increases)
Figure 4.23 Changes of job satisfaction & security led by appraisals

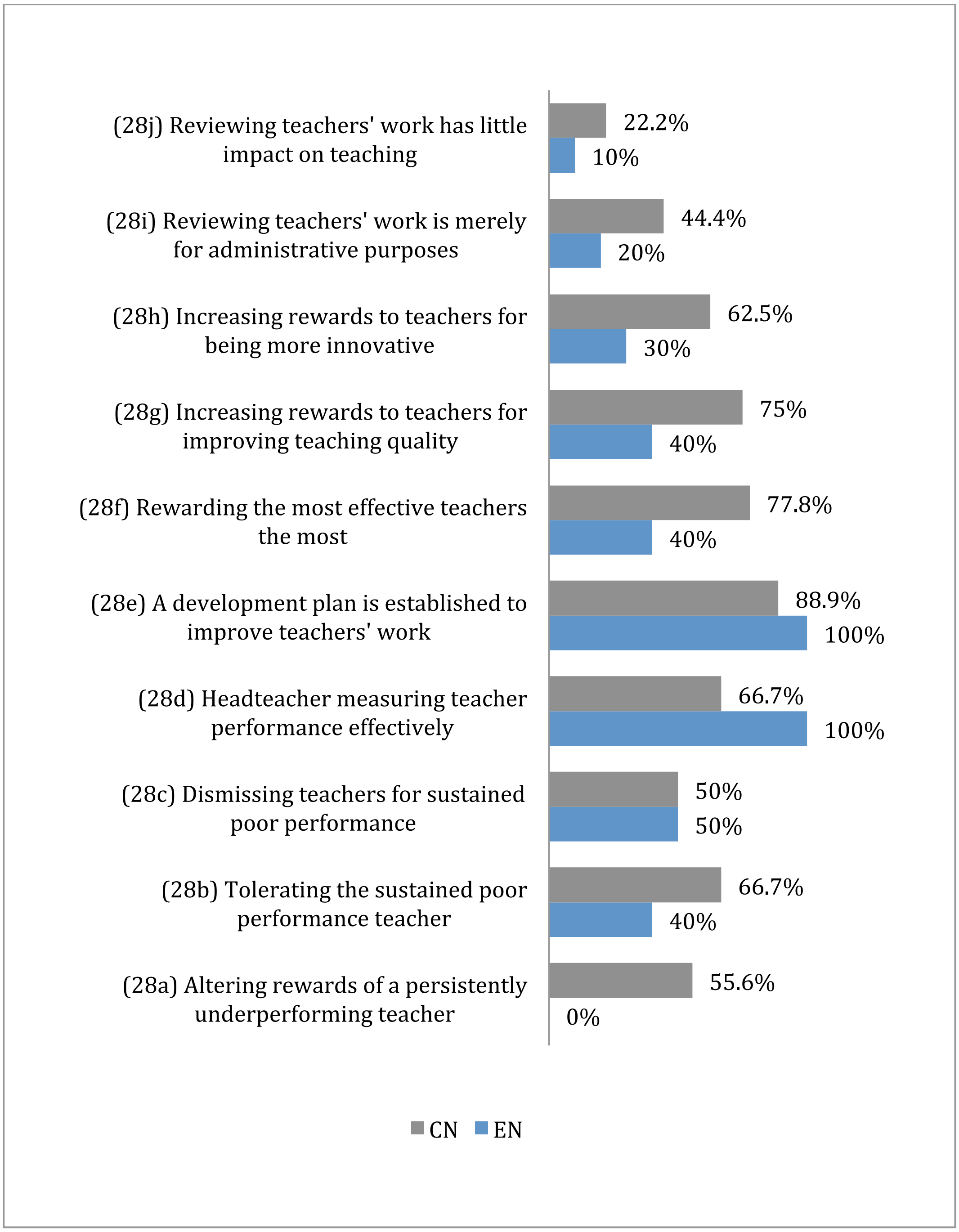
Figure 4.23 suggests that less English teachers perceived appraisal-affected changes in job satisfaction than did Chinese teachers and that less Chinese teachers saw changes in job security than did English teachers.

Teacher perceived school policy on appraisals. As shown in Table 4.15, teachers from two countries perceived the appraisals/feedback in their schools significantly differently in four aspects – #28a, #28b, #28d and #28i (contents may be found in Figure 4.24) – each with a strong effect.

Table 4.15 Teacher perceived school policy on appraisals

#	*Nation*	*1*	*2*	*3*	*4*	X^2	*df*	*p*	*φ*
28a **Δ	CN	0	4	5	0	12.124	2	.002	.799
	EN	7	3	0	0				
28b *Δ	CN	0	3	6	0	6.365	2	.041	.579
	EN	5	1	4	0				
28c	CN	1	3	6	0	.032	2	.984	-
	EN	1	4	5	0				
28d *Δ	CN	0	3	6	0	6.041	2	.049	.564
	EN	0	0	7	3				
28e	CN	0	1	8	0	4.025	2	.134	-
	EN	0	0	7	3				
28f	CN	1	1	6	1	3.089	3	.378	-
	EN	4	2	3	1				
28g	CN	0	2	4	2	5.657	3	.130	-
	EN	5	1	3	1				
28h	CN	1	2	3	2	3.015	3	.389	-
	EN	5	2	2	1				
28i *Δ	CN	0	5	4	0	6.131	2	.047	.568
	EN	5	3	2	0				
28j	CN	1	6	2	0	6.031	3	.110	-
	EN	5	4	0	1				

Notes
1 = Strongly disagree
2 = Disagree
3 = Agree
4 = Strongly agree
* = Significant at the .05 level
** = Significant at the .01 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval



(Agree & Strongly Agree)

Figure 4.24 Teacher-perceived school policy on appraisals

All English teachers disagreed or strongly disagreed that the head teacher had took steps to alter the monetary rewards of a persistently underperforming teacher (#28a), whereas over a half of Chinese teachers held the opposite opinion. About two thirds of Chinese thought their colleagues would tolerate a teacher's sustained poor performance (#28b), and about the same proportion of English teachers indicated that their colleagues would not. All English teachers trusted that their heads were able to evaluate teaching (#28d), but this proportion dropped to two thirds when it came to the Chinese. The proportion of Chinese teachers holding negative view on the appraisal and/or feedback (#28*i* and #28j) doubled that of English teachers. It is not clear whether this was the truth or it was because Chinese teachers were more critical than their English colleagues. As shown in Figure 4.24, about two thirds of or more Chinese teachers indicated that their schools gave rewards to teachers for being effective (#28f), improving their teaching (#28g) and being innovative (#28h) and that their schools had a development or training plan for their improvement (#28e). Almost the same or more proportion of English teachers disagreed or strongly disagreed to the statements of #28f, #28g and #28h, and all of them agreed or strongly agreed to #28e.

Summary. This section provides a complex picture of teachers' views regarding the appraisal/feedback. All teachers indicated that the appraisal/feedback provided judgement about their work quality and improvement suggestions. There were similar proportions of teachers from both countries perceiving a change in job satisfaction as a result of appraisals, but less Chinese teachers felt their job security improved because of appraisals. Whilst all English teachers thought the appraisal was fair and helpful, so did only 9/10 to 3/4 of Chinese teachers. About school policies on the appraisal/feedback, English teachers held more positive views than did Chinese teachers, whereas Chinese teachers seemed to give a mixture of positive and negative views. Chinese schools were more likely to reward more effective and innovative teachers and alter rewards of persistently poor performance teacher.

Now, we will move on to the next section to look at the results of the survey on teaching practice, beliefs and attitudes.

4.3.4 Teaching practice, beliefs and attitudes

In this section, five aspects of teacher perceptions are presented: (1) *teacher beliefs about teaching and learning*, (2) *school-based professional activities*, (3) *teacher the self and the school*, (4) *school management* and (5) *main subjects that teachers teach*.

4.3.4.1 Teacher beliefs about teaching and learning

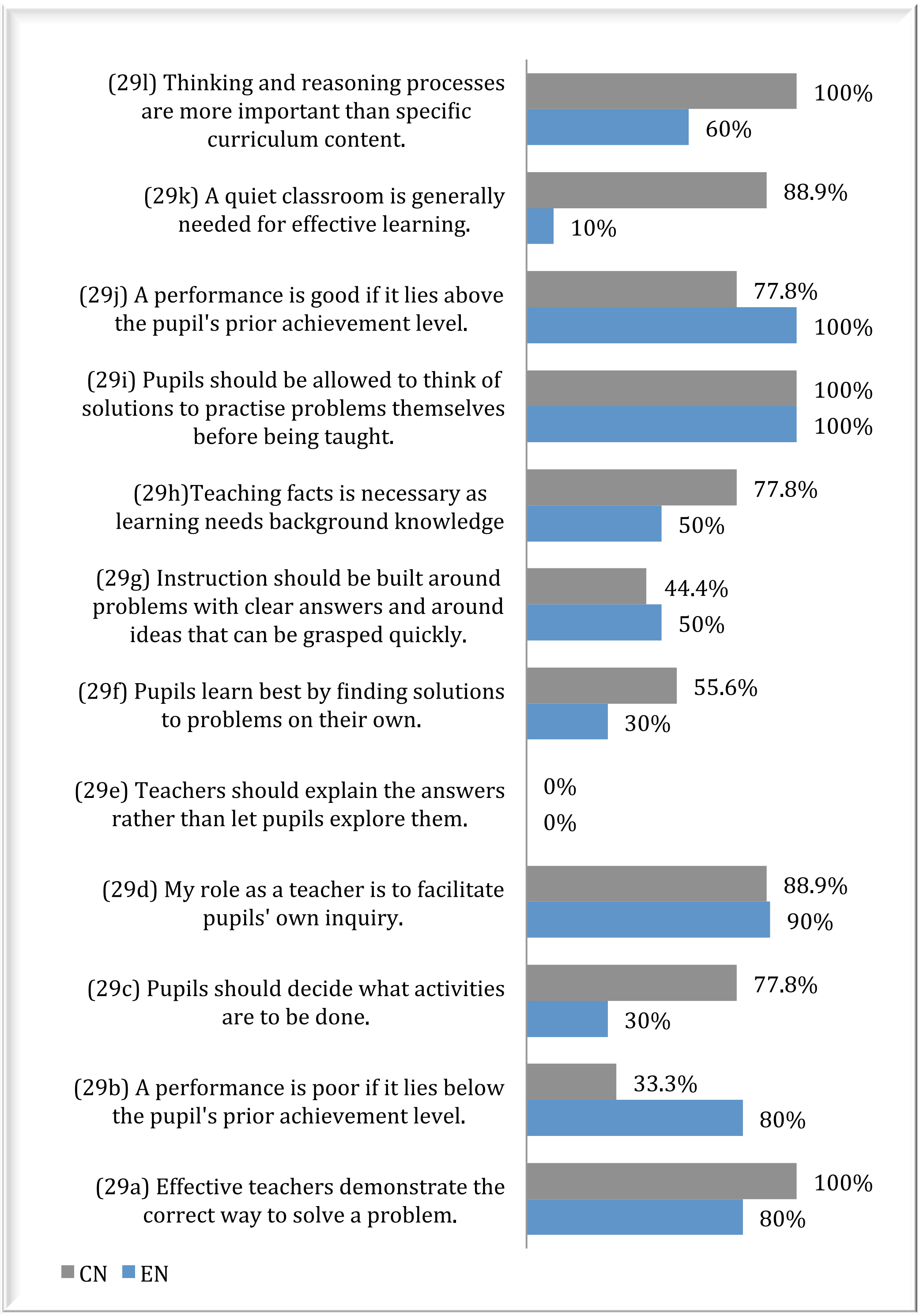
This question (#29) was intended to find out whether teachers held "direct transmission beliefs" or "constructivist beliefs" about teaching and learning (OECD, 2009a, p. 92). Both were surveyed with four items (four-point scales ranging from strongly disagree to strongly agree). Comparing detailed ratings by teachers between two countries, there is a significant difference found in the aspect #29k – *a quiet classroom is generally needed for effective learning* – at the level of .01 with a strong effect; a significant difference is also found in the aspect #29*l* – *thinking and reasoning processes are more important than specific curriculum content* – at the level of .05 with a strong effect (Table 4.16).

Table 4.16 Teacher beliefs about teaching and learning

#	Type	Nation	1	2	3	4	X^2	df	p	φ
29a	DT	CN	0	0	7	2	3.442	2	.179	-
		EN	0	2	4	4				
29b	DT	CN	1	5	3	0	4.847	3	.183	-
		EN	0	2	7	1				
29c	DT	CN	0	2	7	0	3.600	1	.058	-
		EN	0	6	3	0				
29d	C	CN	0	1	8	0	2.020	2	.364	-
		EN	0	1	7	2				
29e	DT	CN	2	7	0	0	.693	1	.405	-
		EN	4	6	0	0				
29f	C	CN	0	4	5	0	3.778	2	.151	-
		EN	0	2	4	3				
29g	DT	CN	1	4	4	0	1.111	2	.574	-
		EN	0	4	5	0				
29h	DT	CN	0	2	7	0	1.571	1	.210	-
		EN	0	5	5	0				
29i	C	CN	0	0	9	0	3.206	1	.073	-
		EN	0	0	7	3				
29j	DT	CN	0	2	6	1	3.958	2	.138	-
		EN	0	0	10	0				
29k **Δ	DT	CN	0	1	6	2	11.996	3	.007	.795
		EN	1	8	1	0				
29l *Δ	C	CN	0	0	5	4	8.061	2	.018	.651
		EN	0	4	6	0				

Notes

1 = Strongly disagree
2 = Disagree
3 = Agree
4 = Strongly agree
* = Significant at the .05 level
** = Significant at the .01 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval
C = Constructivist beliefs
DT = Direct transmission beliefs



(Agree & Strongly Agree)

Figure 4.25 Teacher beliefs on teaching and learning

The general stereotype is that Chinese teachers must fall in the direct-transmission category, and the English should absolutely be constructivist. However, the picture in Figure 4.25 seems much more complex than that. There were indeed three direct transmission aspects (#29a, #29h and #29k) agreed or strongly agreed by a higher percentage of teachers in China than in England where however 50% or 80% of the teacher participants also agreed to #29a or #29h; the aspect #29g seemed agreeable to a similar proportion of Chinese and English teachers. Nonetheless, in the four constructivist aspects, the proportion of Chinese teachers that voted positively for #29f and #29*l* doubled that of English teachers, similar percentages of teachers from both countries achieved an agreement on the aspect #29d, and all teachers agreed on the aspect #29*i*. Overall, it seems that the English mind was not necessarily direct-transmission free and that the Chinese mind tended to be more pro-constructivist than the English.

Why such a surprising finding? Is it because the aforementioned stereotype was simply based on bias or because the existing educational literature is mainly drawn from data from the West or because the world is changing and we are just not conscious enough? It is not urgent to answer the question now, because what teachers believe or claim they believe and/or do might also contradict what they actually do in the classroom. Further connections between teacher beliefs and practices may be found in other sections in this chapter and the section 5.4 in Chapter 5.

4.3.4.2 School-based professional activities

This question asked teachers to rate frequencies of twelve types of activities going on in their schools. From Table 4.17, before dipping into the statistical figures, the inter-country difference can already be seen almost everywhere – an activity frequency skewing to the left in one country tends to skew to the right in another, except for #30e, #30*i* and #30k (please see detailed content of each scale in Figure 4.26). Significant differences were found in five categories, #30a, #30b, #30d, #30j and #30*l*, at the level of either .01 or .05 with a strong or very strong effect.

To compare the country means, a t-test has also been conducted, and results are shown in Figure 4.26. The international difference is significant in all but two types of activities (#30e and #30k). It suggests how different teachers are organised within each country and how school-level policies might indirectly contribute to the differences of learning outcomes between countries. In seven categories of activities (#30a, #30b, #30c, #30d, #30f, #30g and #30*l*), English teachers participated significantly more frequently than their Chinese colleagues. This is reasonable if we take each country's context into consideration. China is less democratic and has greater power-distance than the case of England, hence less involvement of teachers into school management and decision making, such as the activities #30a and #30f. England doesn't unify textbooks and teaching materials at all levels, whereas China, although in a trend of decentralisation, only allows the diversity of textbooks at the provincial level, i.e. within each province textbooks are still unified. This explains the fact that English teachers tended to take part in activities #30b, #30c and #30d more frequently than Chinese teachers who did not have to do so because everything seemed already decided before hand. Inter-discipline is largely a rare fashion in schools in China; also, Chinese primary teachers are specialists who often just focus on one specific subject; English primary teachers are generalists who teach almost every core subject in a class. These three facts make Chinese teachers less likely to coordinate cross-subject activities than their English counterparts.

Table 4.17 Frequencies of school-based professional activities

#	Nation	1	2	3	4	5	6	X^2	df	p	φ
30a *∆	CN	0	2	3	1	0	3	9.975	4	.041	.725
	EN	0	0	0	0	1	9				
30b *∆∆	CN	4	2	1	1	0	0	12.600	5	.027	.837
	EN	0	0	1	5	1	3				
30c	CN	6	0	2	1	0	0	8.905	4	.064	-
	EN	1	1	1	3	0	3				
30d **∆∆	CN	6	0	1	1	0	0	14.993	4	.005	.939
	EN	0	0	0	1	1	7				
30e	CN	0	0	2	0	1	6	5.248	4	.263	-
	EN	1	1	0	0	0	8				
30f	CN	0	1	2	4	2	0	6.075	4	.194	-
	EN	0	0	0	5	2	3				
30g	CN	1	2	1	1	3	1	6.766	5	.239	-
	EN	0	0	0	3	3	4				
30h	CN	0	0	1	2	2	4	10.978	5	.052	-
	EN	3	3	2	1	1	0				
30i	CN	0	0	0	1	5	3	5.262	4	.261	-
	EN	1	1	0	4	3	1				
30j **∆∆	CN	0	0	0	0	5	4	15.657	4	.004	.908
	EN	0	2	2	5	1	0				
30k	CN	0	1	3	2	1	2	5.630	5	.344	-
	EN	2	0	3	4	1	0				
30l **∆∆	CN	3	4	0	2	0	0	14.989	4	.005	.888
	EN	0	0	0	2	3	5				

Notes

1 = Never
2 = Less than once per year
3 = Once per year
4 = 3-4 times per year
5 = Monthly
6 = Weekly

* = Significant at the .05 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval
** = Significant at the .01 level
ΔΔ = Very strong effect (≥ .8).

Into four categories of activities (#30h, #30i, #30j and #30k), Chinese teachers got involved more frequently than English teachers, which made them significantly different from English teachers in the participating frequency of the first three types.

For a Chinese class, there are a number of teachers teaching different subjects and a class tutor (in Chinese 班主任, literally, the class director) who often teaches Chinese or Mathematics and plays a pastoral role. The class tutor is also responsible in coordinating collaborations between different subject teachers. For a primary class in England, there is often just a teacher and a TA who most of the time does what the teacher asks her/him to do to support low-ability or SEN children. As teachers are based in classrooms, English teachers seem more isolated than their Chinese counterparts who share close connections to the same class with a number of colleagues. It is perhaps because of the facts stated above that two countries' teachers rated the frequency of activity #30h considerably differently.

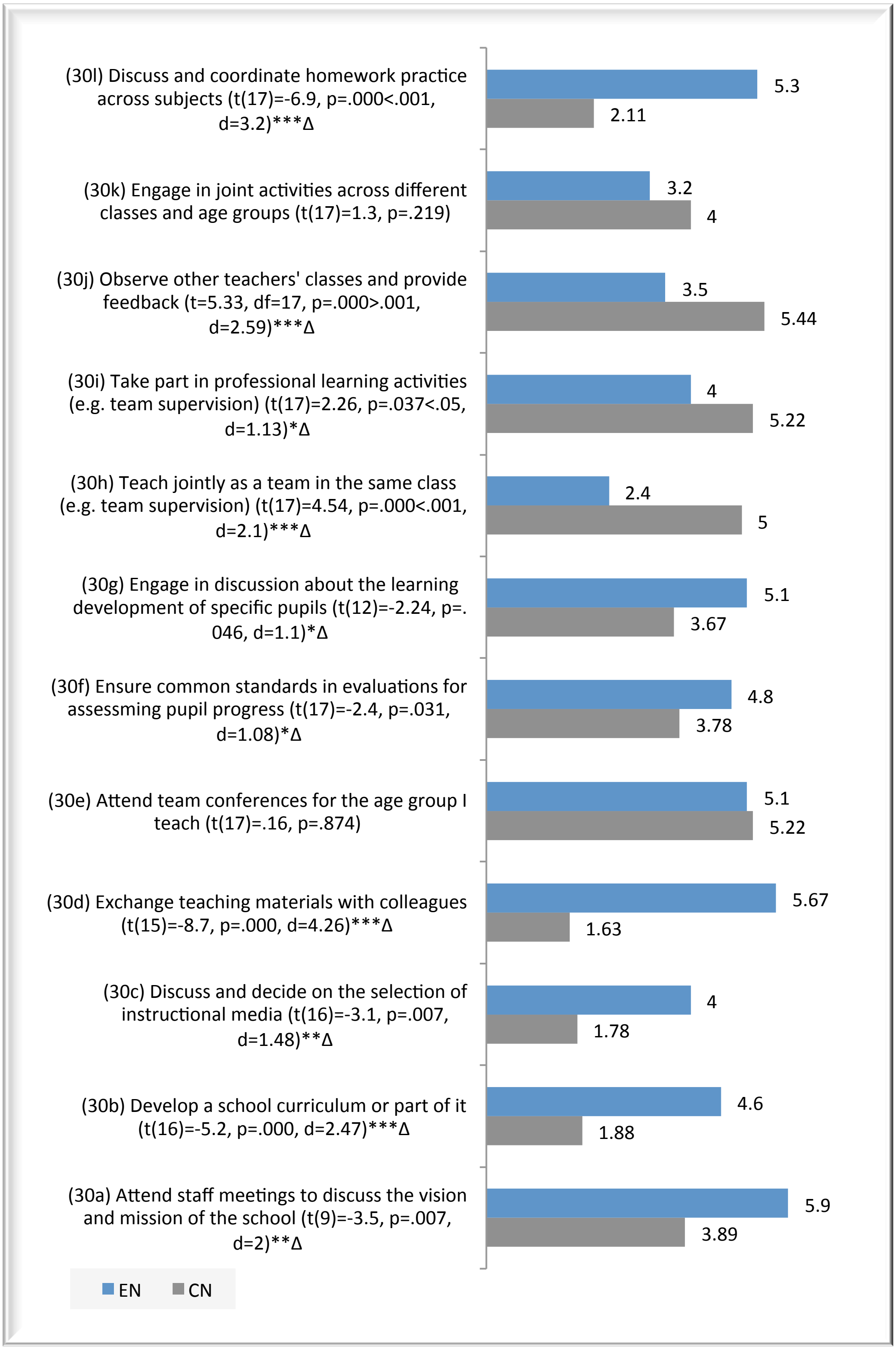


Figure 4.26 Frequencies of school-based professional activities

Note:
1=Never
2=Less than once per year
3=Once per year,
4=3-4 times per year
5=Monthly
6=Weekly.

* = Significant at the .05 level
** = Significant at the .01 level
*** = Significant at the .001 level
Δ = Strong effect (d > 1)

Frequent peer observation and feedback among teachers is also a common phenomenon in Chinese schools; there are also frequent activities in which a group of colleagues (teaching the same subject within and/or across age groups) and internal and/or external leaders observe a lesson together and then jointly discuss the quality of the lesson and possible improving plans immediately afterwards. Within Chinese schools, there are also subject-based and age-group-based teaching research groups in which regular (often weekly) meetings are held for teachers to discuss strategies of improving teaching and learning quality and tackling specific issues arising recently. It is thus not surprising to see the much higher frequencies of #30i and #30j rated by Chinese teachers than by English teachers. Based on similar reasons, the difference in #30k is also understandable. In Chinese schools, teaching seems to be more of a collective wisdom than an individual mission; in English schools, teaching is more of a teacher's individual exploration than a joint effort.

4.3.4.3 Teacher the self and the school

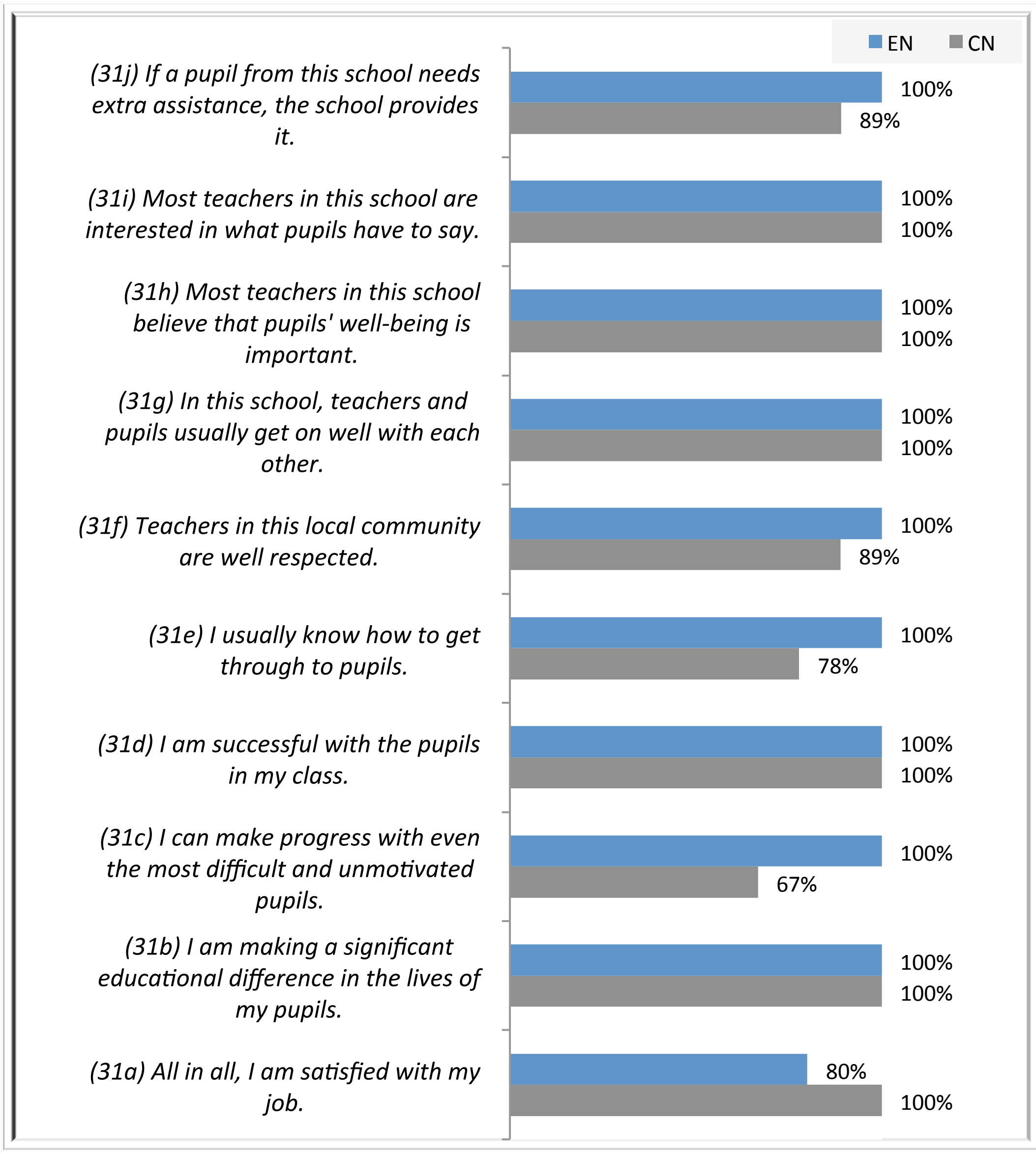
Question #31 is about teachers' perceptions of their own teaching (#31a to #31f) and their school climate (#31g to #31j). As shown in Table 4.18, although four items were rated significantly differently between countries, the difference mainly lied in the degree of agreement. Differences may be seen more clearly through combining *agree* and *strongly agree* percentages together within each country (see Figure 4.27).

Table 4.18 Teacher the self and the school

	#	Nation	1	2	3	4	X^2	df	p	φ
Self	31a	CN	0	0	8	1	5.095	2	.078	-
		EN	0	2	4	4				
	31b	CN	0	0	8	1	3.316	1	.069	-
		EN	0	0	5	5				
	31c	CN	0	3	5	1	5.630	2	.060	-
		EN	0	0	5	5				
	31d	CN	0	0	9	0	7.892	1	.005	.645
	**Δ	EN	0	0	4	6				
	31e	CN	0	2	7	0	6.041	2	.049	.564
	*Δ	EN	0	0	6	4				
	31f	CN	0	1	8	0	4.025	2	.134	-
		EN	0	0	7	3				
School	31g	CN	0	0	6	3	0.540	1	.463	-
		EN	0	0	5	5				
	31h ***	CN	0	0	9	0	19.000	1	.000	1.000
	ΔΔ	EN	0	0	0	10				
	31i	CN	0	0	9	0	9.975	1	.002	.725
	**Δ	EN	0	0	3	7				
	31j	CN	0	1	7	1	5.352	2	.069	-
		EN	0	0	4	6				

Notes

1 = Strongly disagree
2 = Disagree
3 = Agree
4 = Strongly agree
* = Significant at the .05 level
** = Significant at the .01 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval



(Agree & Strongly Agree)

Figure 4.27 Teacher the self and the school

In Figure 4.27, the proportion of English teachers with an overall job satisfaction (#31a) is one fifth less than that of Chinese teachers who, however, are less likely to be satisfied with their ability of dealing with pupils (#31c and #31e) than their English counterparts. A very striking result is that less Chinese teachers felt well respected by the local community (#31f), given the high status that teachers are said to be holding in Confucius-heritage societies. English teachers all agreed that their schools were supportive to SEN pupils (#31j), and so did eight out of nine Chinese teachers. Overall, Chinese teachers seemed to be less confident than their English colleagues.

Next section is about the teacher-perceived frequencies of management behaviours of their headteachers.

Table 4.19 Frequencies of headteachers' management behaviours

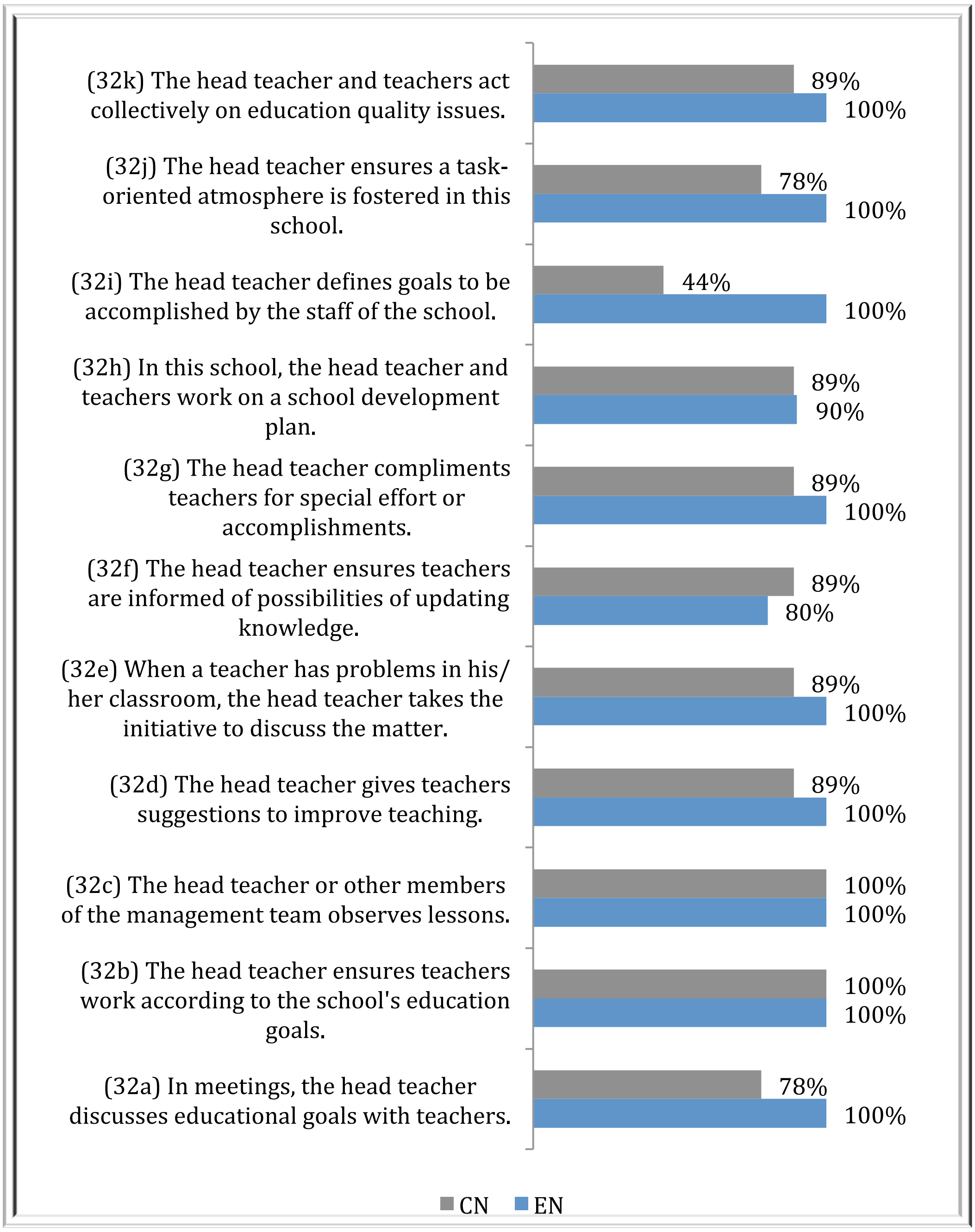
#	*Nation*	*1*	*2*	*3*	*4*	X^2	*df*	*p*	*φ*
32a *Δ	CN	0	2	5	2	6.852	2	.033	.601
	EN	0	0	2	8				
32b *Δ	CN	0	0	6	3	6.537	1	.011	.587
	EN	0	0	1	9				
32c	CN	0	0	3	6	.024	1	.876	-
	EN	0	0	3	7				
32d	CN	0	1	6	2	3.357	2	.187	-
	EN	0	0	4	6				
32e	CN	0	1	6	2	1.619	2	.445	-
	EN	0	0	6	4				
32f	CN	0	1	7	1	4.560	2	.102	-
	EN	0	2	3	5				
32g	CN	0	1	7	1	1.351	2	.509	-
	EN	0	0	8	2				
32h *Δ	CN	0	1	7	1	7.245	2	.027	.618
	EN	0	1	2	7				
32i *Δ	CN	1	4	3	1	9.474	3	.024	.706
	EN	0	0	3	7				
32j	CN	0	2	6	1	4.718	2	.094	-
	EN	0	0	5	5				
32k *Δ	CN	0	1	7	1	7.067	2	.029	.610
	EN	0	0	3	7				

Notes

1 = Never

2 = Seldom

3 = Quite often 4 = Very often
* = Significant at the .05 level
Δ = Strong effect, i.e. fall in the [0.5, 0.8) interval



(*Quite Often* & *Very Often*)

Figure 4.28 Frequencies of headteachers' management behaviours

4.3.4.4 School management

Eleven management behaviours were provided for teachers to rate the corresponding frequencies as carried out by their heads on a scale ranging from *never* to *very often* (Table 4.19). Though significant inter-country differences were found in five behaviours each with a strong effect, the difference in items #32b, #32h and #32k only lies in the degree of agreement, and only #32a and #32i tell a noticeable international difference.

Accumulated proportions of positive responses in Figure 4.28 further show that Chinese teachers seem not quite satisfied with their heads' way of managing the schools and they have less say in school management than English teachers. This echoes with the results in the question #31. However, the item #32i might need a second thought – it could be that Chinese teachers, rather than their heads, are responsible for defining goals for themselves. If so, Chinese teachers might have more autonomy in their own work than the English, though they are less involved in managing the school.

Next, we will look at the last aspect of this questionnaire section which is about the main subjects that teachers teach.

4.3.4.5 Main subjects that teachers teach

All English teachers teach Literacy, Maths and PE, and two thirds or more of them also teach Science, Modern Languages, IT, Arts, Religion/Ethics and other subjects, such as History, Geography and Music (Figure 4.29). Chinese teachers all teach Mathematics, with three of them each teaching either Arts, PE or another subject at the same time. It seems that Chinese teachers have less workload than English teachers, if only judging by the number of subjects they teach.

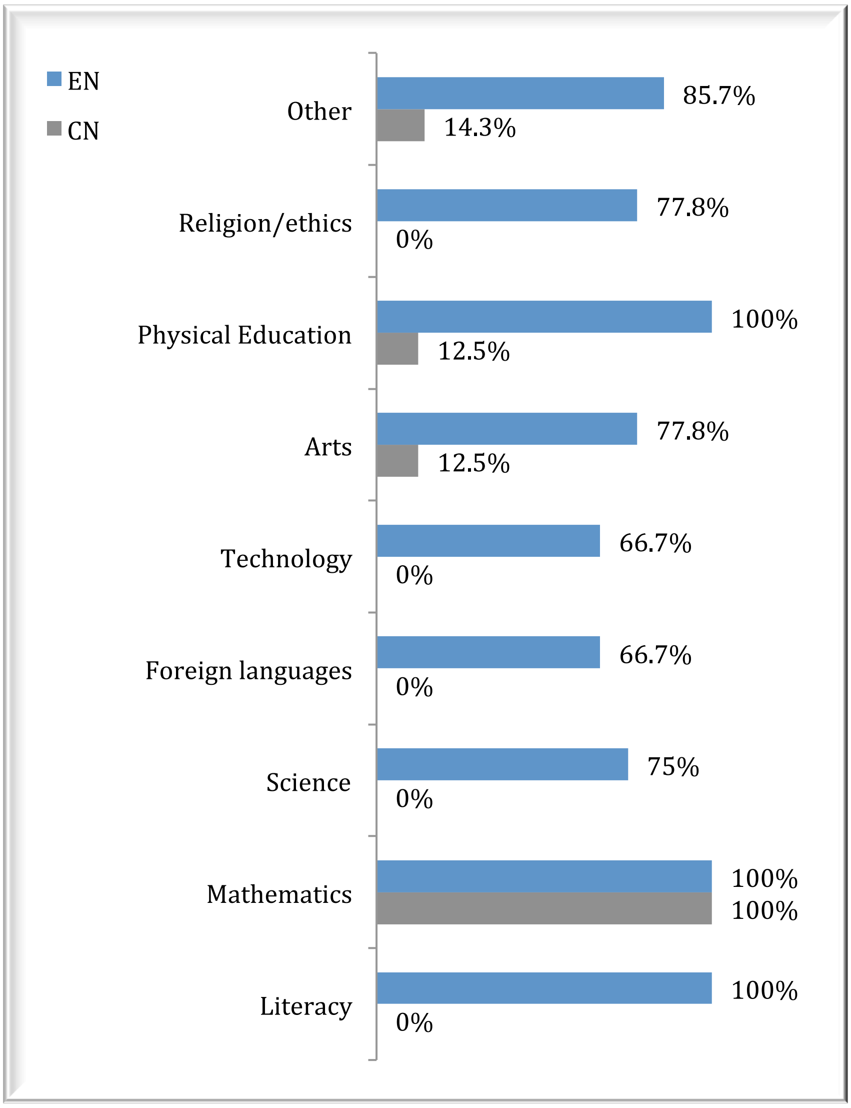


Figure 4.29 The main subjects that teachers teach

Next, we will move on to the final section of the questionnaire to see information about the observed class.

4.3.5 Teaching in the observed class

This section presents the characteristics of the observed class, time and frequency of various activities in the class and its disciplinary climate as judged by teachers.

4.3.5.1 Characteristics of the class

This subsection presents teacher-informed features of the class in three aspects: (1) *class sizes*, (2) *pupil abilities* and (3) *pupil family background*.

Class sizes. Table 4.20 shows that England has much smaller class sizes than China. The international difference is significant, and the relationship between nationalities and class sizes is very strong.

Table 4.20 A comparison of class sizes between England & China

Item	*Nation*	*N*	*Mean*	*SD*	*t*	*df*	*p*	*d*
34 ***Δ	CN	8	37.5	4.690	6.164	16	.000	2.8
	EN	10	26.6	2.757				

Notes:
*** = Significant at the .001 level   Δ = Strong effect ($d > 1$)

Pupil abilities. Teachers were asked to rate their pupils' abilities comparing to other pupils in the same grade/year level in the same school (#35a) and to pupils in the same grade/year level more generally (#35b). Results in Table 4.21 suggest that Chinese classes had a considerably smaller spread of pupil abilities than English classes and that none of Chinese teachers perceived the existence of any low ability children in their classes. This might be partially due to the fact that English primary schools group children by ability, whereas Chinese primary schools take the mixed-ability approach.

Table 4.21 The ability of pupils in the teacher's class

#	Nation	1	2	3	4	5	X^2	df	p
35a	CN	0	0	5	3	0	5.070	4	.280
	EN	1	1	4	1	2			
35b	CN	0	0	5	3	0	3.453	4	.485
	EN	1	1	3	3	1			

Notes
1 = Much lower than average ability
2 = Slightly lower than average ability
3 = Average ability
4 = Slightly higher than average ability
5 = Much higher than average ability

4.3.5.2 **Time and frequency of various activities**

Teachers were asked to estimate the amounts of time that they spent in the class on administrative tasks (#37a), keeping order in the classroom (#37b) and actual teaching and learning (#37c). Figure 4.30 shows the comparison of time arrangements across classrooms.

At the country level, English teachers, on average, spent longer time on administrative task and keeping order than Chinese teachers, and the international difference of class time on the former type of activities was significant with a strong effect. It is worth noting that Chinese teachers allocated 4.85% more class time on actual teaching and learning than did English teachers, despite the lack of a statistical significance (Table 4.22).

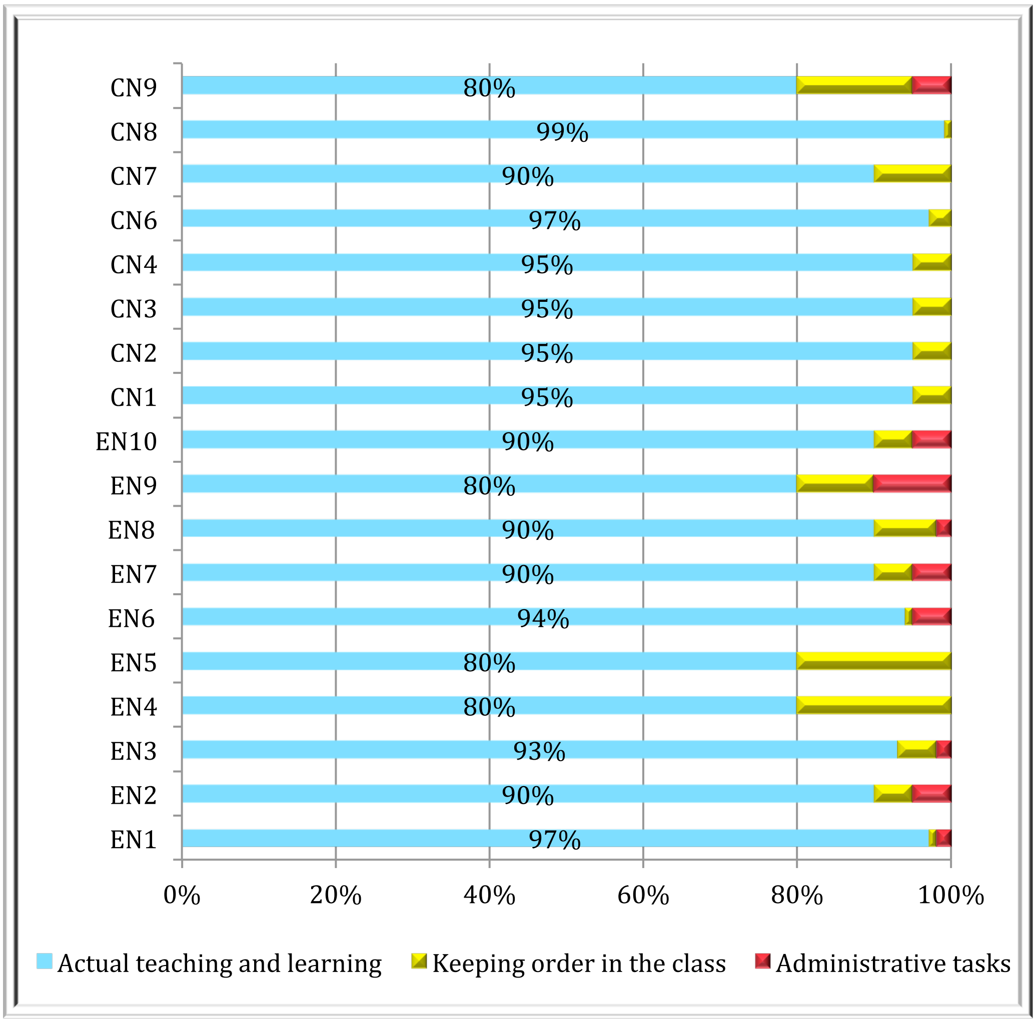


Figure 4.30 A comparison of time allocations across classrooms

Table 4.22 An international comparison of class time allocations

Items	*Nation*	*N*	*Mean*	*SD*	*t*	*df*	*p*	*d*
37a *Δ	CN	8	.63	1.768	2.457	16	.026	1.24
	EN	10	3.60	3.026				
37b	CN	8	6.13	4.390	-.668	16	.514	-
	EN	10	8.00	6.880				
37c	CN	8	93.25	5.922	1.678	16	.113	-
	EN	10	88.40	6.222				

Note:
* = Significant at the .05 level   Δ = Strong effect (*d* > 1)

In addition to the arrangement of class time, teachers were also asked to rate the frequencies of given activities over a school year. Table 4.23 suggests that English teachers carried out more frequently five activities in their lessons over a school year, including explicitly stating learning goals (#38b), differentiation by ability (#38e), checking pupils' exercise books (#38i), pupils evaluating or reflecting upon their own work (#38l), and pupils working in groups based upon abilities (#38n). It should be pointed out that English teachers had zero standard deviations across their ratings on the first four activities. These five lesson activities took place at considerably different frequencies in two countries' classrooms.

4.3.5.3 The disciplinary climate of the class

The last question for teachers was about an essential classroom environment indicator – pupil disciplines. Items #39a, #39c and #39d were statements of negative characteristics of pupils, and item #39c positive. Cross tabulation and chi square tests did not show international differences in any aspects (Table 4.24). Figure 4.31 combines percentages of the *Agree* and *Strongly Agree* choices teachers made within each country. It shows that more English classrooms had noise (#39d), whereas more Chinese teachers had to wait for a long time for pupils to quiet down before lessons (#39a) and/or lost lots of time because of pupils' interruption (#39c). More proportion of English teachers found their pupils willing to create a pleasant learning atmosphere (#39b).

Table 4.23 Frequencies of specific lesson activities over the year

Items	*Nation*	*N*	*Mean*	*SD*	*t*	*df*	*p*	*d*
38a	CN	8	3.75	.886	2.000	16	.063	-
	EN	10	2.70	1.252				
38b *Δ	CN	7	3.43	1.512	-2.750	6	.033	2.08
	EN	10	5.00	.000				
38c	CN	8	2.50	1.069	.713	15	.487	-
	EN	9	2.11	1.167				
38d	CN	8	2.38	1.061	-1.840	16	.084	-
	EN	10	3.30	1.059				
38e ***Δ	CN	8	2.13	1.246	-6.524	7	.000	4.61
	EN	10	5.00	.000				
38f	CN	8	1.38	.744	- .646	16	.527	-
	EN	10	1.70	1.252				
38g	CN	8	3.25	1.165	-1.514	16	.149	-
	EN	10	4.10	1.197				
38h	CN	8	4.25	1.165	.102	16	.920	-
	EN	10	4.20	.919				
38i *Δ	CN	8	3.38	1.408	-3.265	7	.014	2.3
	EN	10	5.00	.000				
38j	CN	8	1.25	.463	-1.122	16	.279	-
	EN	10	1.80	1.317				
38k	CN	8	4.13	1.356	- .767	16	.454	-
	EN	10	4.60	1.265				
38l **Δ	CN	8	3.25	1.389	-3.564	7	.009	2.52
	EN	10	5.00	.000				
38m	CN	8	4.75	.707	-1.000	7	.351	-
	EN	10	5.00	.000				
38n ***Δ	CN	8	1.50	1.414	-6.624	15	.000	3.54
	EN	9	4.78	.441				
38o	CN	8	1.25	.707	.603	16	.555	-
	EN	10	1.10	.316				
38p	CN	8	3.13	.991	2.007	16	.062	-
	EN	10	2.40	.516				
38q	CN	8	1.13	.354	.158	16	.876	-
	EN	10	1.10	.316				
38r	CN	8	3.50	1.414	- .627	16	.539	-
	EN	10	3.90	1.287				
38s	CN	8	2.50	1.604	1.641	9.174	.135	-
	EN	10	1.50	.707				

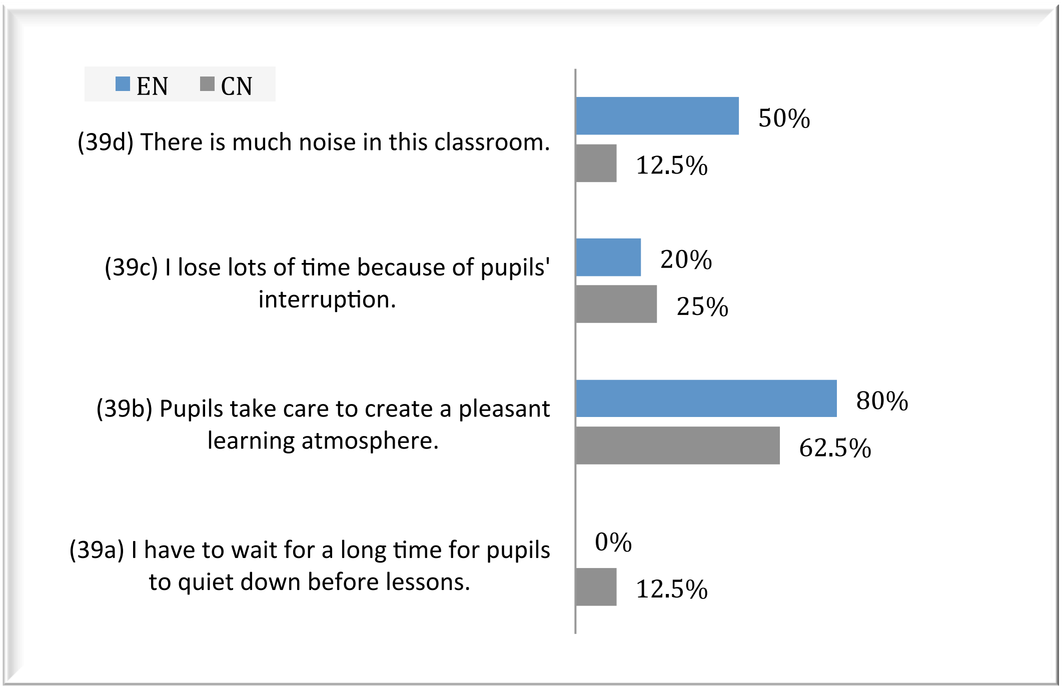
Note:

* = Significant at the .05 level
** = Significant at the .01 level
*** = Significant at the .001 level
Δ = Strong effect (*d* > 1)

Table 4.24 Pupil disciplines in the class as perceived by the teacher

#	Nation	1	2	3	4	X^2	df	p
39a	CN	2	5	1	0	1.613	2	.446
	EN	4	5	0	0			
39b	CN	0	3	4	1	2.822	2	.244
	EN	0	2	3	5			
39c	CN	1	5	1	1	6.043	3	.110
	EN	6	2	2	0			
39d	CN	1	6	1	0	3.825	2	.148
	EN	2	3	5	0			

Note:
1 = Strongly disagree
2 = Disagree
3 = Agree
4 = Strongly agree



(*Agree* & *Strongly Agree* to the statements)

Figure 4.31 Classroom environment & pupil behaviours

Overall, English teachers seemed to be more optimistic about classroom disciplines than the Chinese, apart from their noisy classrooms for which some English teachers applauded, with written reasons attached, such as “noise is good!”

4.4 Pupil questionnaire (M2.2)

In this section, two strands of PQ data are presented to provide information about *pupil-perceived schooling* and *pupil-perceived maths learning and teaching*. The remaining part of PQ data is interpreted in section 4.7 together with three TQ items as pupil background information.

4.4.1 Pupil-perceived schooling

This sub-section focuses on issues relating to children’s schooling, including their perceptions of their prior performance, their time in the current class, the frequency of homework and the proportion of homework they usually complete.

Perceived previous performance. Regarding their self-esteems, pupils were asked, “When you are younger, what kind of marks (or grades) did you usually get in school?” As shown in Table 4.25, there seem to be no difference between two countries if we only compare the sum of *Very high* to *Good* pupils of each country. Nonetheless, the total rates of the *Very high* and *High* rose 10% higher in China than in England. Interestingly, on the lower end of the scale, about 2% more children from China felt they were previously not very good at mathematics, comparing with the corresponding proportion of English children. The international difference regarding detailed choices on this topic was significant at the level of .01 with a modest effect.

Table 4.25 Child-perceived previous performance of the subject

PQ #73		*5*	*4*	*3*	*2*	*1*	*Total*	*X^2*	*df*	*p*	*φ*
CN	N	58	115	48	99	18	338	13.518	4	.009	.154
	%	17.2%	34.0%	14.2%	29.3%	5.3%	100%				
EN	N	36	60	58	68	7	229				
	%	15.7%	26.2%	25.3%	29.7%	3.1%	100%				
Total	N	94	175	106	167	25	567				
	%	16.6%	30.9%	18.7%	29.5%	4.4%	100%				

Note:
5 = Very high 4 = High 3 = Good 2 = Some good, some not 1 = Not very good

Table 4.26 Children's time in their current classes

PQ #74		*1*	*2*	*3*	*4*	*Total*	*X^2*	*df*	*p*	*φ*
CN	N	0	6	3	329	338	8.407	3	.038	.122
	%	0%	1.8%	0.9%	97.3%	100%				
EN	N	1	2	9	215	227				
	%	0.4%	0.9%	4%	94.7%	100%				
Total	N	1	8	12	544	565				
	%	0.2%	1.4%	2.1%	96.3%	100%				

Note:
1 = For less than two weeks 3 = Between one and two months
2 = For about three to four weeks 4 = More than two months

Time in this class. The majority of children in either country had been in the class where they took part in the study for more than two months (Table 4.26), which means they were all in a relatively familiar class. The rates went higher for the English regarding the proportions of children joining the class within 1-2 weeks or 1-2 months. The international difference was significant at the level of 0.05 with a modest effect. The result may suggest that pupils in England are slightly more likely to travel across schools than pupils in China where the *hukou* (户口) policy sets more obstacles preventing people from moving from city to city and region to region within the country.

Amount of time per week on homework. The majority of English children reported that they spent either half an hour or one hour per week on homework, whereas the majority of Chinese pupils spent 3-4 or 5-7 hours during a normal school week (Table 4.27). Two countries were significantly different in this respect at the level of .001 with a strong effect.

Table 4.27 Children's time spent on homework per week

PQ #64		*A*	*B*	*C*	*D*	*E*	*F*	*G*	*Total*
CN	N	10	3	14	48	132	100	34	341
	%	2.9%	0.9%	4.1%	14.1%	38.7%	29.3%	10%	100%
EN	N	10	108	69	17	17	8	1	230
	%	4.3%	47%	30%	7.4%	7.4%	3.5%	0.4%	100%
Total	N	20	111	83	65	149	108	35	571
	%	3.5%	19.4%	14.5%	11.4%	26.1%	18.9%	6.1%	100%

Note:
A = No time
B = Half an hour
C = 1 hour
D = 2 hours
E = 3-4 hours
F = 5-7 hours
G = more than 8 hours
X^2 = 340.071
df = 6
p < .001
φ = .772

Days per week having homework. Slightly over a half of English pupils said they had homework one day per week; almost 99 per cent of Chinese pupils reported that they had homework every working day (Table 4.28). The international difference was significant at the level of .001 with a strong effect.

Table 4.28 Days per week having homework

PQ #65		*1 day*	*2 days*	*3 days*	*4 days*	*5 days*	*Never assigned*	*Total*
CN	N	0	1	1	0	337	2	341
	%	0.0%	0.3%	0.3%	0.0%	98.8%	0.6%	100%
EN	N	131	12	11	13	61	2	230
	%	57.0%	5.2%	4.8%	5.7%	26.5%	0.9%	100%
Total	N	131	13	12	13	398	4	571
	%	22.9%	2.3%	2.1%	2.3%	69.7%	0.7%	100%

Note:
X^2 = 344.478 *df* = 5 *p* < .001 *φ* = .777

Table 4.29 Children's homework completion rates

PQ #63		*A*	*B*	*C*	*D*	*E*	*F*	*Total*
CN	N	1	2	16	38	137	147	341
	%	0.3%	0.6%	4.7%	11.1%	40.2%	43.1%	100%
EN	N	0	4	22	31	137	36	230
	%	0.0%	1.7%	9.6%	13.5%	59.6%	15.7%	100%
Total	N	1	6	38	69	274	183	571
	%	0.2%	1.1%	6.7%	12.1%	48%	32%	100%

Note:
A = Never assigned
B = None of it
C = Some of it
D = Most of it
E = All
F = All plus some extra
X^2 = 51.001
df = 5
p < .001
φ = .299

Completion rates of homework. Slightly over a half of English pupils expressed that they usually did all the homework assigned, and so did slightly less than a half of Chinese pupils (Table 4.29). This does not mean that Chinese children were less likely to complete homework assigned, because about the same proportion of them did not only all but also extra homework. Combining the proportions of children completing "All" and "All plus some extra" homework, we get the ratio of England to China which is 73.1 : 83.3. The international difference regarding homework completion rates was significant at the level of .001 with a modest effect.

Summary. About previous performance, Chinese pupils seemed slightly more confident than their English buddies. The majority of all pupils have been in their current classes for over two months, but a noticeable fact is that the proportion of the English children joining the class within two months doubled the corresponding Chinese proportion. Chinese pupils had more homework on almost every school day and were also more likely to complete their homework than English pupils.

4.4.2 Pupil-perceived maths learning and teaching

As noted in section 3.4.2 only 34 PQ items were clearly classified as the 7 C's in the available MET publication. This section is thus only built on the results of the 34 items. Each of these items measures pupil perceptions on a five-point scale (1

= No/Never, 2 = Mostly Not, 3 = Maybe/Sometimes, 4 = Mostly Yes, and 5 = Yes/Always).

Table 4.30 An international comparison of pupil perceptions

7 C's [full score]	Nation	*N*	*Mean*	*SD*	*t*	*df*	*p*	*d*
Care [35] ***+	CN	326	28.09	5.87	-4.14	550.1	0.000	0.38
	EN	227	29.93	3.92				
Control [20] ***+	CN	330	16.01	2.77	4.97	517.0	0.000	0.43
	EN	220	14.91	2.36				
Clarify [40]	CN	328	33.79	5.03	-0.51	546.8	0.611	--
	EN	223	33.98	3.64				
Challenge [10] *** ++	CN	338	7.30	2.14	-8.50	561.4	0.000	0.71
	EN	230	8.54	1.33				
Captivate [20]	CN	336	16.57	3.53	0.71	548.5	0.480	--
	EN	228	16.39	2.79				
Confer [35] ** +	CN	325	28.77	4.48	-2.70	544.3	0.007	0.23
	EN	226	29.68	3.41				
Consolidate [10] *** Δ	CN	338	6.15	2.04	-13.35	547.8	0.000	1.13
	EN	228	8.22	1.63				

Note:
** = Significant at the 0.01 level
*** = Significant at the .001 level
+ = Modest effect
++ = Moderate effect
Δ = Strong effect

Pupils from both countries perceived *Captivate* (items #21, #30, #34 and #58) and *Clarify* (items #6, #9, #14, #19, #23, #41, #47 and #51) at a similar level, which suggests no international differences (Table 4.30). English pupils perceived a higher degree of teaching in four dimensions: *Care* (items #4, #7, #16, #20, #33, #52 and #57), *Challenge* (items #8 and #22), *confer* (items #11, #26, #31, #35, #38, #43 and #50) and *Consolidate* (items #32 and #39) than their Chinese peers

($p < .001$). Chinese pupils perceived significantly more *Control* (items #10, #12, #13 and #55) than English pupils ($p < .001$).

The next two parts of this chapter serve to present detailed results of the two mathematics tests that children took and the correlations between teaching and learning.

4.5 Standardised mathematics tests (M3)

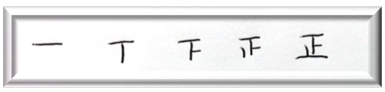
In this section, test results are presented in four ways to generate insight from different angles. This results in breaking the section into four subsections:

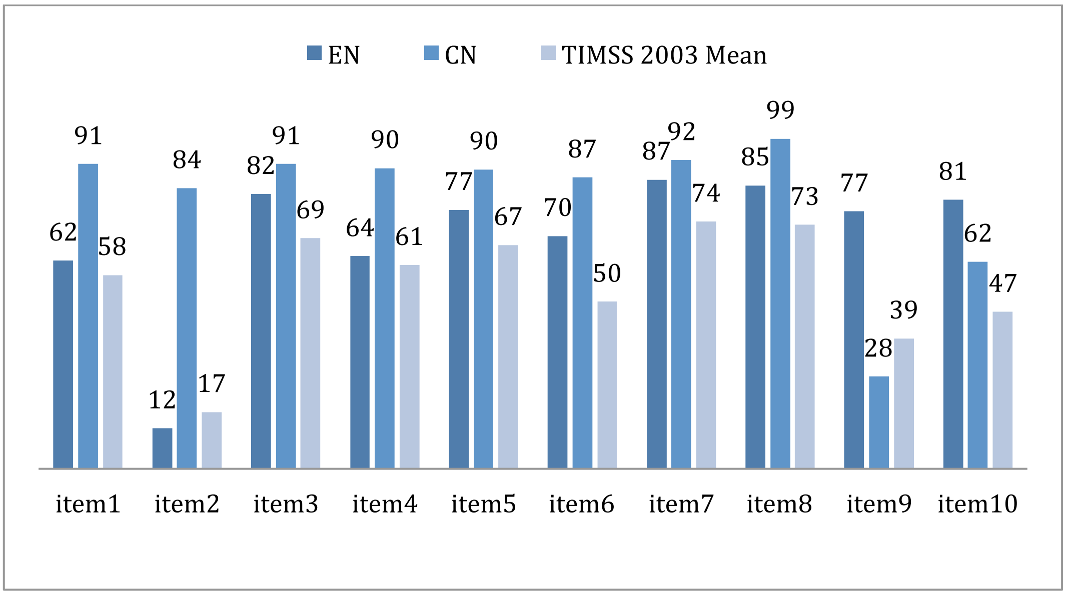
- Test results by item
- Test results by domain
- Overall test results
- Correlating learning outcomes with teaching

4.5.1 Test results by item

Within each country, the correct rate of each item was calculated through dividing the number of children who did the item correctly by the total number of all that took the test. Test one saw the correct-rate gaps of 29 items between England (lower rates) and China (higher) ranging from 18 (inclusive) to 72 per cent and such international gaps on 7 items ranging from 1 to 14 per cent (Figure 4.32). While both countries had the same rates – 86 per cent – on items #17, England outperformed China on items #9, #10 and #13 by 49%, 19% and 17% respectively. It should be pointed out that the items #9 and #10 represented numbers with the western style of tally marks. This might have caused the relatively lower correct rates of Chinese pupils who were used to the Chinese equivalent style – writing the five strokes of the Chinese character 正[4]. Comparing with the TIMSS 2003 average, English correct rates flew higher on 21 items, with the difference ranging from one to 38 per cent, but English rates were also two to 35 per cent lower in the remaining 19 items. Under the same comparison, Chinese correct rates exceeded the TIMSS average by 18 to 67 per cent on 31 items and 7 to 16 per cent on 7 items and lagged behind the TIMSS 2003 mean by 10 and 11 per cent on items #9 and #10 respectively which, as aforementioned, involved the representation of numbers with the western-style tally marks.

In test two, English correct rates were 43, 18, 10, 1 and 1 per cent higher than the Chinese ones on five items - #9, #10, #13, #17 and #22 respectively; Chinese correct rates were higher than the English ones on all remaining 35 items; the Chinese-English gap ranged from 18 to 62 per cent on 22 items, and from 4 to 17 per cent on 13 items (Figure 4.33).

4 



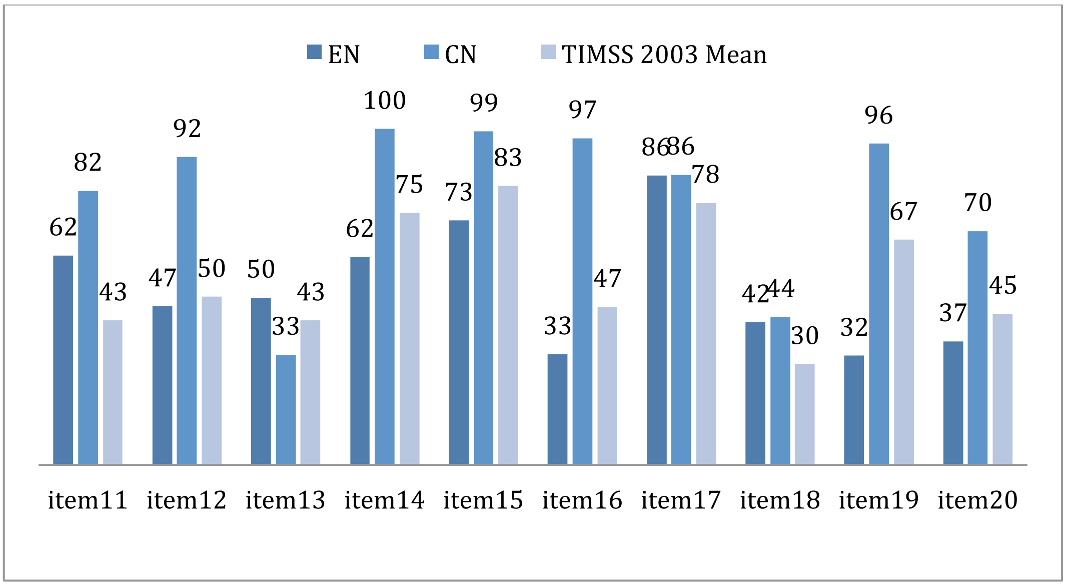
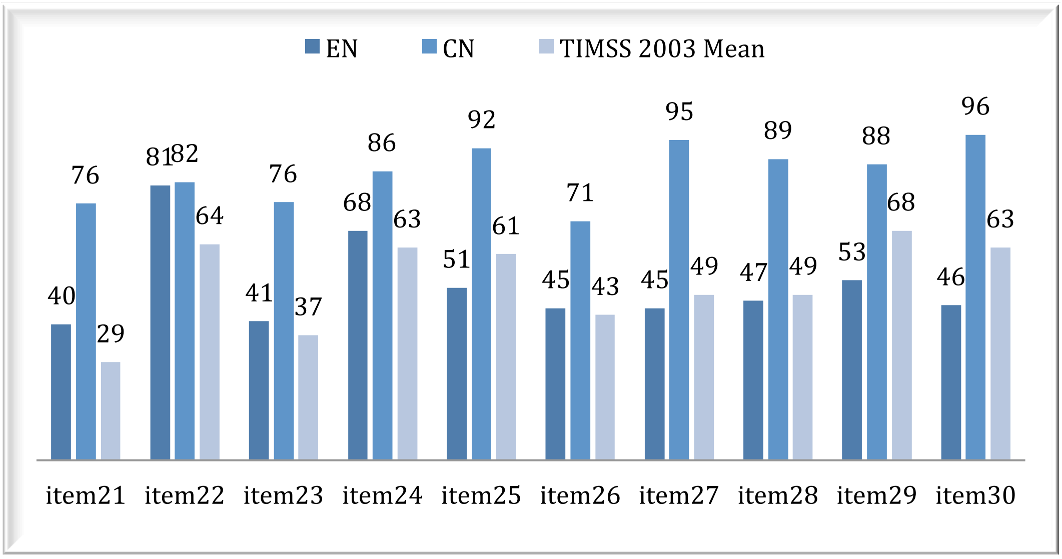
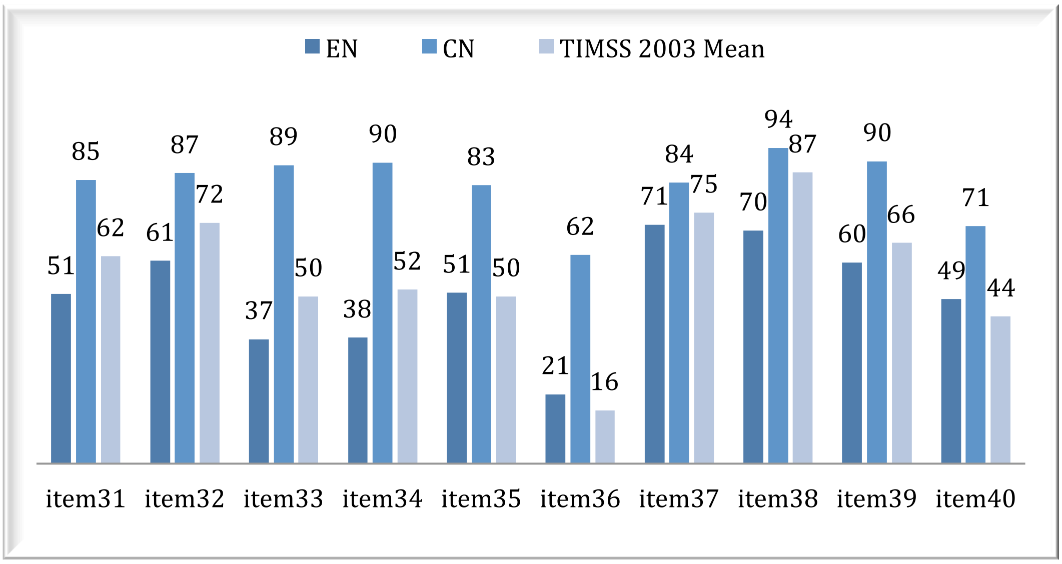


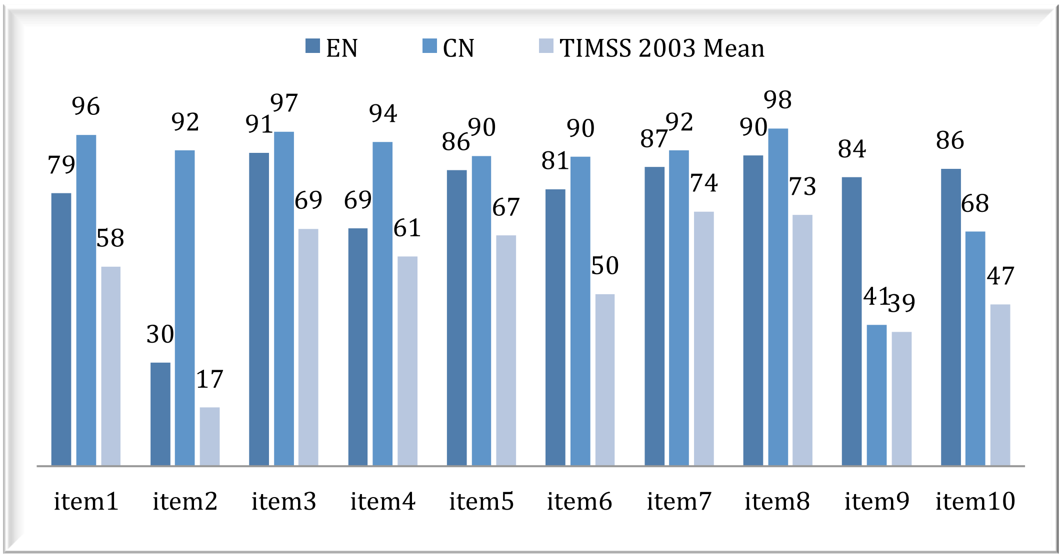
Figure 4.32 A comparison of item correct rates in Test 1 (%)

(Figure 4.32 continuing to the next page)





(End of Figure 4.32)



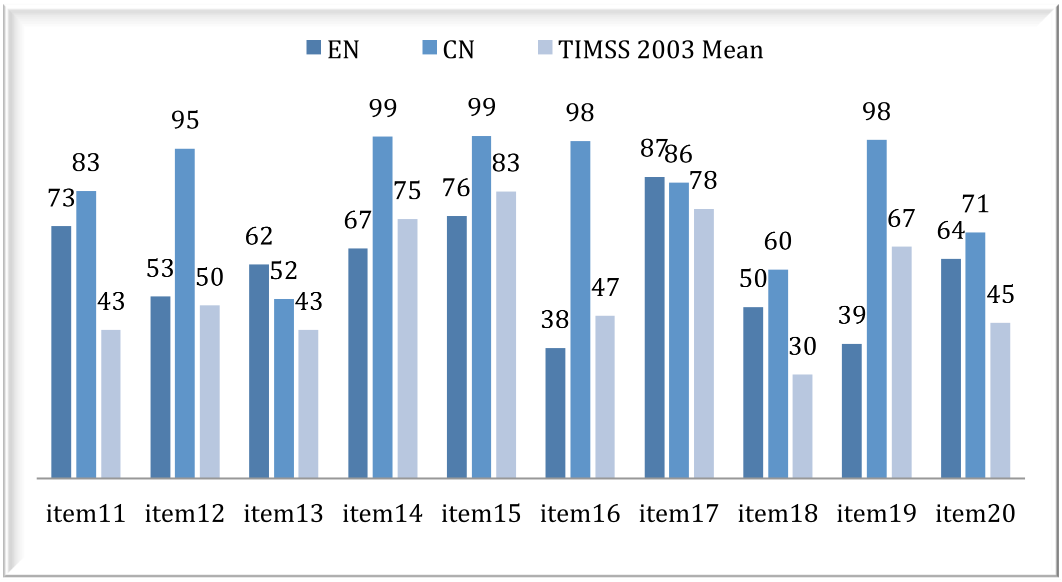
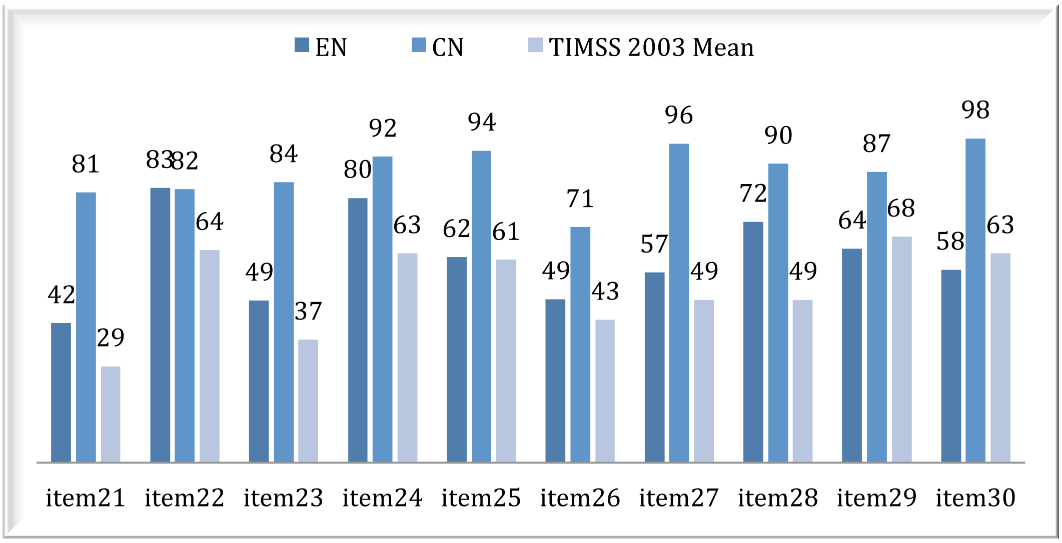
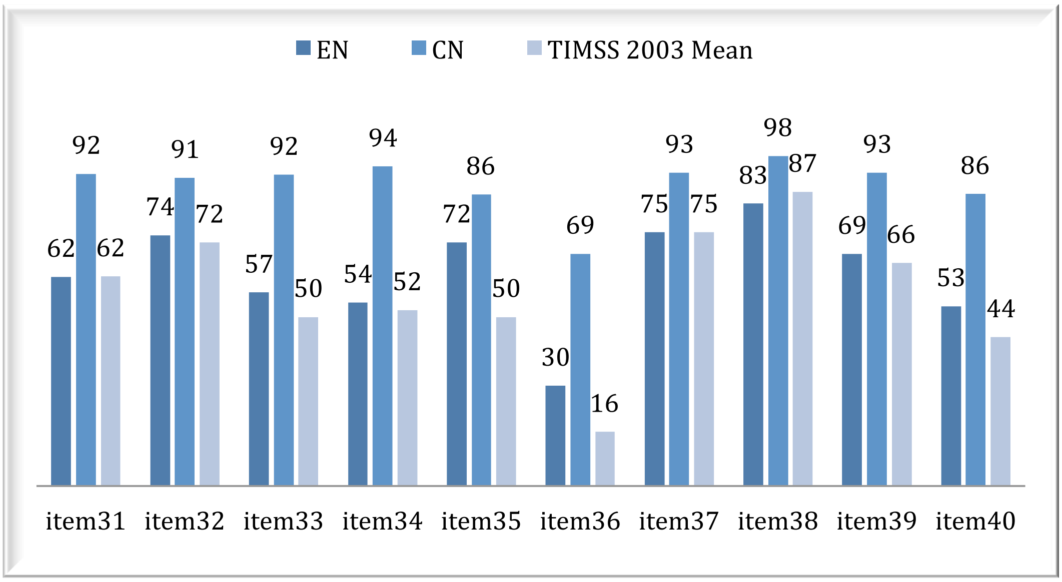


Figure 4.33 A comparison of item correct rates in Test 2 (%)

(Figure 4.33 continuing to the next page)





(End of Figure 4.33)

Vertically (or historically) comparing with the TIMSS 2003 average, English pupils had achieved 20 to 45 per cent higher correct rates on nine items, 12 to 19 per cent higher rates on 11 items, one to 9 per cent higher rates on 11 items; they got four to 28 per cent lower rates on seven items and the same rates as the

TIMSS average on two items – #31 and #37. Following the same comparison route, Chinese pupils achieved higher rates on all items of which they made 18 to 75 per cent higher correct rates on 35 items than the TIMSS average, 8 to 16 per cent higher on 4 items, and two per cent higher on item #9.

Overall, Chinese pupils made much higher correct rates in comparison with either English pupils in this study or pupils on the average line in TIMSS 2003. The positive differences between the correct rates of England and TIMSS 2003 average, where there were, tended to spread evenly across various ranges, whereas such gaps between Chinese correct rates and the TIMSS 2003 average skewed dramatically to the higher end.

Having seen the results of the two tests by item, next, we will look at the results by content domain and cognitive domain.

4.5.2 Test results by domain

As shown in Chapter 3, each TIMSS item belongs to a specific content domain and a particular cognitive domain. This subsection compares the correct rate of each content domain and cognitive domain cross-nationally.

In test one, in terms of correct rates, English pupils' best content domain was *Data* (Figure 4.34), and their best cognitive domains were *Solving routine* *problems* and *Using concepts* (Figure 4.35). Chinese pupils' strongest content domain was *Patterns and relationship*, and their best cognitive domain was *Knowing facts and procedures*.

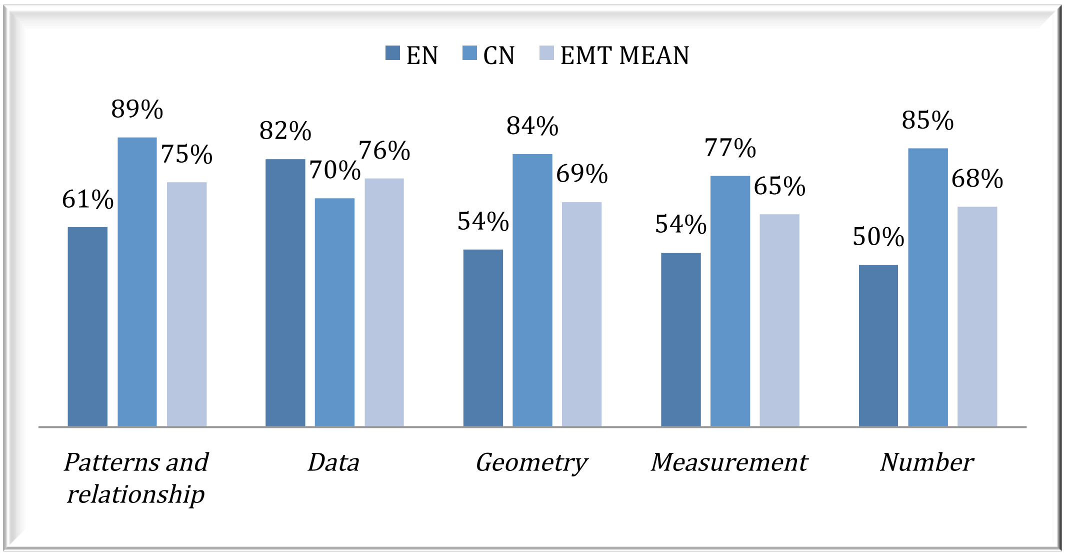


Figure 4.34 Correct rates by content domain in Test 1

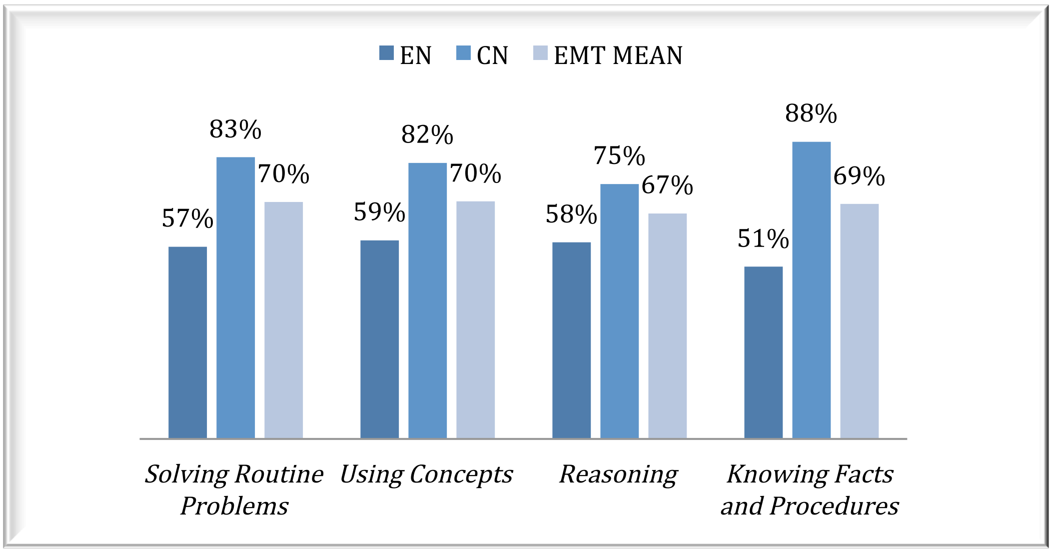


Figure 4.35 Correct rates by cognitive domain in Test 1

In test one, English pupils achieved the least correct rate in *Number* and *Knowing facts and procedures*; Chinese pupils had the lowest rate in *Data* and *Reasoning*. English pupils correct rates were all below the EMT international average but the content domain of *Data* in which they exceeded the average by 6 per cent and the Chinese rate by 12 per cent. Apart from *Data*, English pupils were below the EMT average in any other content or cognitive domain by 9 to 19 per cent and lower than Chinese rates by 23 to 36 per cent in four other content domains and by 17 to 37 per cent in all four cognitive domains; Chinese pupils were above the EMT average by 12 to 19 per cent in four other content domains and all four cognitive domains.

In test two (Figure 4.36 & Figure 4.37), Chinese pupils again got the lowest correct rate in the domain of *Data* but much higher rates in all other domains. English pupils' highest correct rate was in the two domains of *Data* and *Reasoning,* in comparison with their performance in other domains. Chinese pupils made the least mistakes in the domains of *Patterns and relationship*, *Using concepts* and *Knowing facts and procedures* among all domains. Again, except for *Data*, English pupils attained 6 to 15 per cent lower correct rates than the EMT average and 13 to 30 per cent lower than their Chinese peers in all other four content and four cognitive domains; Chinese pupils got 6 to 15 per cent higher correct rates than the EMT average in all other domains.

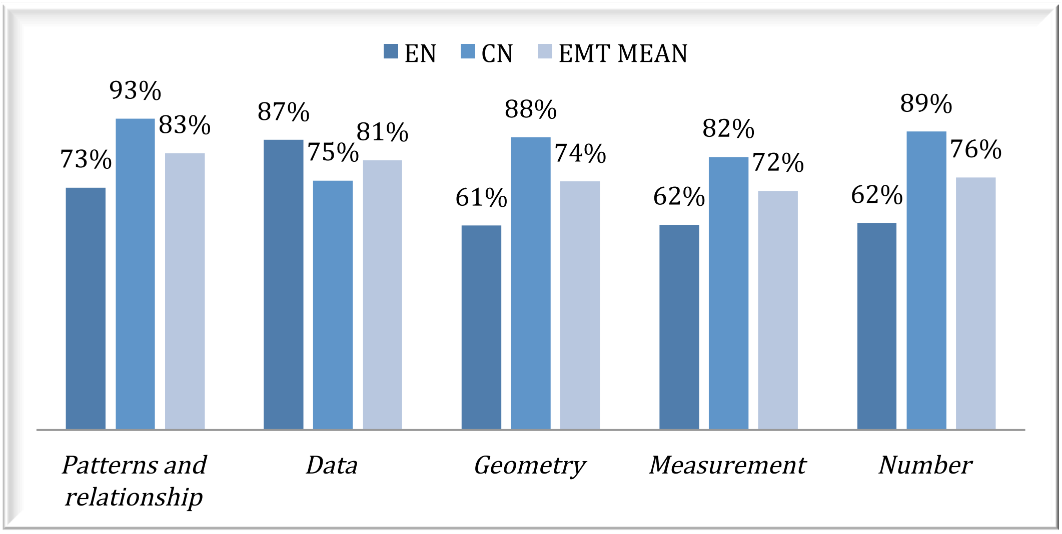


Figure 4.36 Correct rates by content domain in Test 2

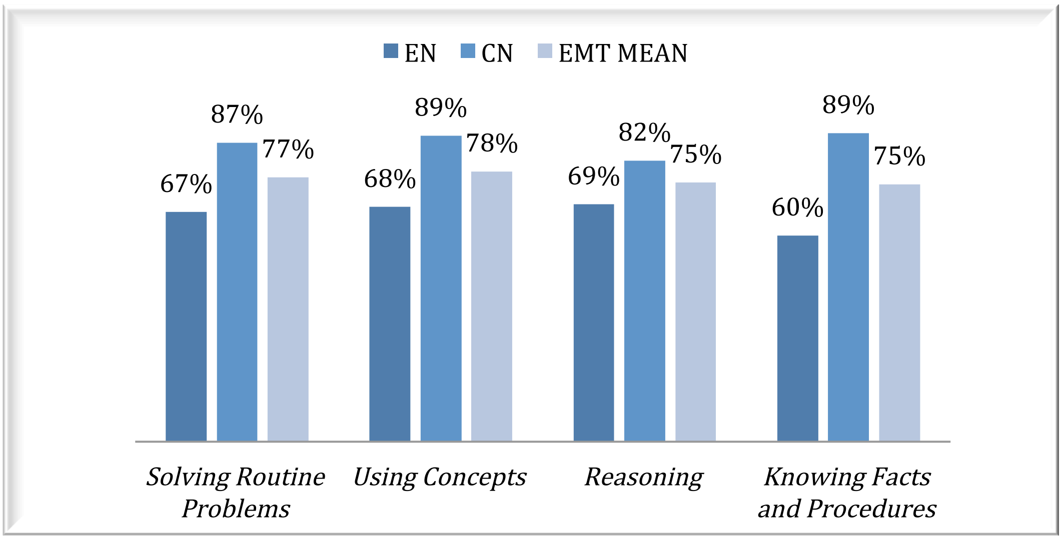


Figure 4.37 Correct rates by cognitive domain in Test 2

Having scrutinised children's performance at the item and domain levels, we will now zoom out to see the full picture – the overall test results.

4.5.3 Overall test results

Both the two maths tests showed Chinese pupils on average outperformed their English peers by over 20 per cent. As follows, detailed test results will be presented, and the ceiling effect upon Chinese pupils will be explained.

Test one results

In test one, with a difference of 27, English and Chinese pupils had achieved a mean of 56 and 83 respectively. The within country variance was quite different (England: N = 231, SD = 21; China: N = 250, SD = 10). The international difference was significant and the relationship between nationality and pupil performance was strong ($t(321) = -17.7$, $p < .001$, $d = 1.74 > 1$).

Test two results

In test two, on average, the English pupils have achieved 66 (N = 236, SD = 19), and the Chinese 87 (N = 326, SD = 10), which indicated a gap of 21 between two countries on average. The former have gained approximately 10 percent, probably due to a quite low starting point and therefore sufficient space for improvement; the latter only made it to a 4 percent increase. The international difference was statistically significant and strong ($t(325) = -15.4$, $p < .001$, $d = 1.44 > 1$).

Ceiling effect and learning gains

English pupils had made a progress of 9.7 over the ten school weeks between two tests, whereas their peers in China only improved 4.2 on average. This does not mean that Chinese teachers had added less value on to pupils' learning in mathematics over time than had their English colleagues. The Chinese means in two tests were both about 10 points higher than the corresponding international means, whereas English means were over 10 points lower than the average on both occasions. It was assumed that there might be a ceiling effect on Chinese pupils in the second test that did not allow their scores to grow. This assumption was tested through the following measures.

To find the proportions of pupils that had achieved a score that was close to or reached the ceiling, a line through the 60 percentile score was drawn to see how many pupils from each country had made it to the top 40% in the two tests. In test one, the test score separating top 40% achievers from the rest was 80 (i.e. the 60 percentile score); in test two, the score was 86.3. Then, further calculations were made regarding the proportions of English and Chinese pupils that were among the top 40%, i.e., between the 60 percentile (80 for test one and 86.3 for test two) and the ceiling (100). Results are shown in Figure 4.38. In either test, none of English pupils had made it to the ceiling, but a small proportion of Chinese pupils had already achieved a full mark. In test one, 13.4% of English pupils and 42.7% of Chinese pupils were amongst the top 40% in this cohort of 9- to 10-year olds. In test two, the proportions were 12.9% (EN) and 40.3% (CN) respectively. There is indeed a ceiling effect upon Chinese pupils which did not provide enough space for them to improve their scores between two tests.

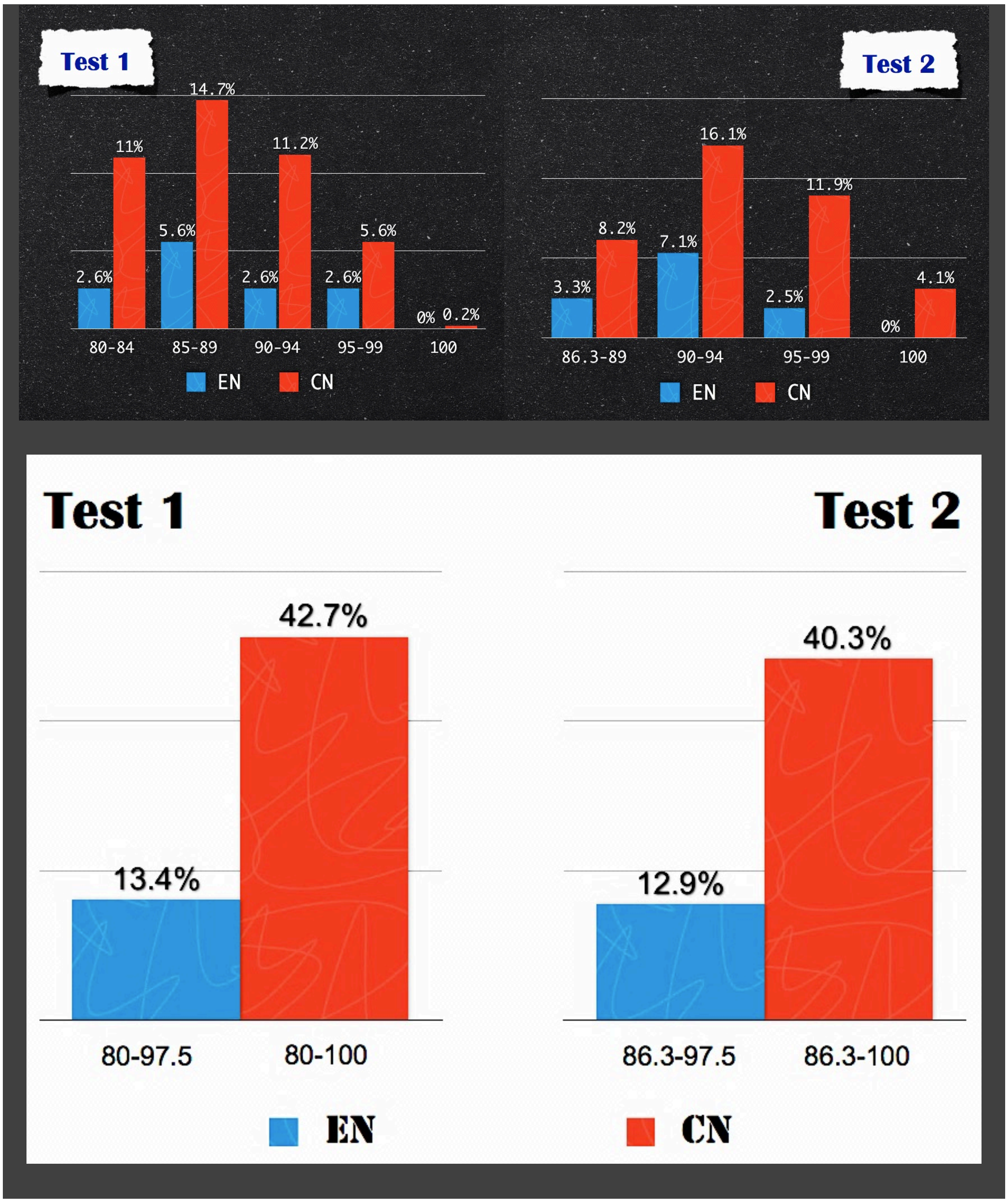


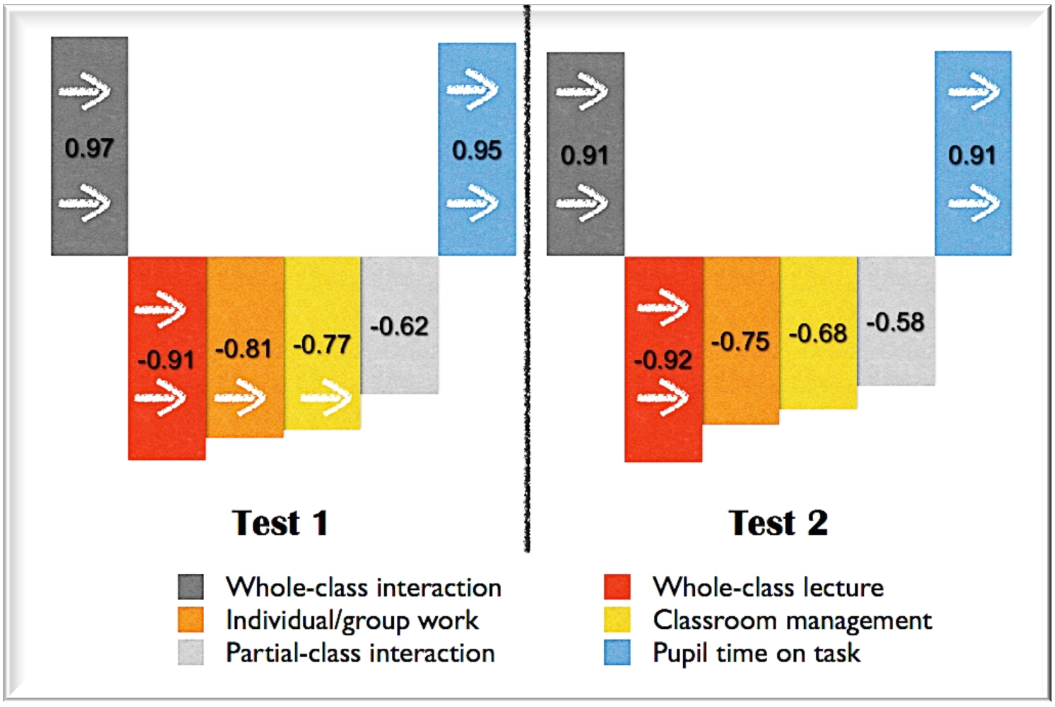
Figure 4.38 The proportions of top 40% pupils in two tests

This ceiling effect might be caused by the limit of the TIMSS 2003 items which were too easy for Chinese pupils. These items were apparently not suitable for testing learning gains in this study. It was thus decided that teacher behaviours should be correlated with pupils' maths performance at points in time, rather than learning gains over time.

Now that we have seen both teaching and learning data, it's time to explore the correlations between them.

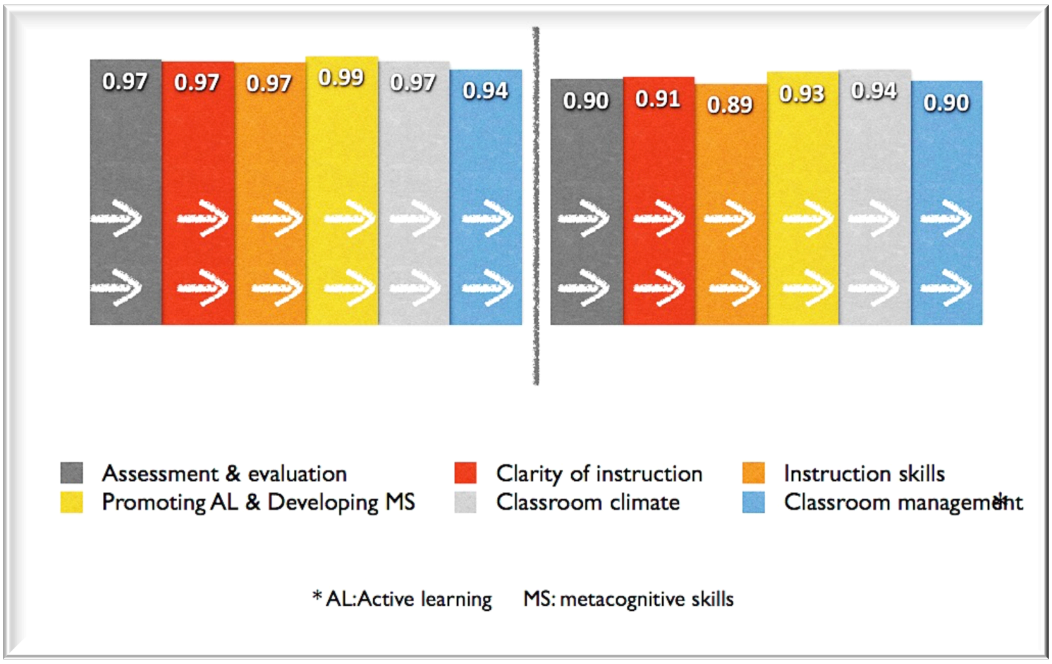
4.6 Correlations between maths teaching & learning

Teacher scores in OTL and ISTOF were correlated with pupil performance in each test at the classroom level across two countries (Figure 4.39 and Figure 4.40).



Note: → = significant at the .05 level →→ = significant at the .01 level

Figure 4.39 Correlations between OTL and learning outcomes



Note: →→ = significant at the .01 level

Figure 4.40 Correlations between ISTOF and learning outcomes

The pooled analysis suggested that eight teacher behaviours/factors were positively correlated with mathematics performance and three others had a negative impact. These effective teaching behaviours/factors are:

- Whole-class interaction
- Pupil time on task
- Assessment and evaluation
- Clarity of instruction
- Instructional skills
- Promoting active learning & developing metacognitive skills
- Classroom climate
- Classroom management (quality)

Those negative and ineffective teaching behaviours/factors are:

- Whole-class lecture
- Individual/group work
- Classroom management (quantity/frequency)

Partial-class interaction might also have posed negative effects upon pupil learning, but the impact was not statistically significant.

4.7 Pupil background differences

This section is mainly built on partial data collected with the pupil questionnaire, including children's basic information, such as gender and ethnicity, and home background information, such as the number of computers, frequencies of speaking English at home, number of children, number of adults and number of books in the child's bedroom. Data collected with three teacher-questionnaire items (#36a, #36b, #36c) are also included to show parents' highest levels of education as perceived by teachers. The purpose of compiling this section is to try and understand, with the limited scope of data, whether there are any differences of pupil intakes between two countries and, if so, how likely such differences may affect maths performance internationally.

4.7.1 Gender and ethnicity

There were a slightly higher proportion (approx. 6% more) of boys in the Chinese cohort than in the English (Table 4.31). This might be a result of the fact that traditionally and historically Chinese favoured boys over girls. Although it is widely perceived that this gender bias is not an issue at least in major cities nowadays, the result does suggest that such a bias was still in the mind of a small number of parents in Nanjing at least around the year 2003 when these pupils were born. Internationally, however, the gender ratio was not significantly different.

Table 4.31 Proportions of boys and girls in the two countries

PQ #66		*Boy*	*Girl*	*Total*	X^2	*df*	*p*
CN	N	190	148	338	1.697	1	.193
	%	56.2%	43.8%	100%			
EN	N	116	113	229			
	%	50.7%	49.3%	100%			
Total	N	306	261	567			
	%	54%	46%	100%			

Table 4.32 An overview of English children's ethnic background

PQ #72		***WB***	***WO***	***IN***	***P***	***M***	***B***	*Total*
EN only	N	188	14	4	0	14	1	228
	%	82.5%	6.1%	1.8%	0%	6.1%	0.4%	100%
		C	***A***	***BC***	***BO***	***O***		
	N	2	1	0	1	3		
	%	0.9%	0.4%	0%	0.4%	1.3%		

Note:
WB = White British
WO = White (other)
IN = Indian
P = Pakistani
M = Mixed
B = Bangladeshi
C = Chinese
A = Other Asian (non Chinese)
BC = Black Caribbean
BO = Black others
O = Other

Ethnicity. In the English sample, about four fifths of pupils were White British, and one fifth were from minority background (Table 4.32). The proportion of ethnic minority dropped significantly in the Chinese cohort (Table 4.33).

Table 4.33 An overview of Chinese children's ethnic background

PQ #72		***Han***	***Minority***	*Total*
CN ***only***	N %	327 97%	10 3%	337 100%

4.7.2 Number of computers at home

On the one hand, the proportion of English pupils who did not have a computer at home was higher than that of Chinese pupils in the same circumstance; on the other hand, the proportion of Chinese pupils who had either one or more computers was higher than that of their English counterparts (Table 4.34). These seemingly small differences were found statistically significant at the level of .01 with a modest effect.

4.7.3 Speaking English at home

This question was only for English pupils. The result in Table 4.35 seems to coincide with the proportion of White British in Table 4.32.

Table 4.34 Number of computers per household

PQ #67		*No*	*Yes (=1)*	*Yes (>1)*	*Total*	X^2	*df*	*p*	*φ*
CN	N	8	138	190	336	9.513	2	.009	.13
	%	2.4%	41.1%	56.5%	100%				
EN	N	18	86	123	227				
	%	7.9%	37.9%	54.2%	100%				
Total	N	26	224	313	563				
	%	4.6%	39.8%	55.6%	100%				

Table 4.35 Frequency of speaking English at home

PQ #68		*Yes, always*	*Yes, sometimes*	*No, never or almost never*	*Total*
EN **only**	N	195	22	10	227
	%	85.9%	9.7%	4.4%	100.0%

4.7.4 **Number of children at home**

Children were asked, "Counting yourself and all others, how many children live with you?" The most rated number for the English was "two", whereas that for the Chinese was unsurprisingly "one" because of the one-child policy in China. The actual percentage might be higher for China, as some children usually took their cousins as their siblings given the strong family bond among Chinese. The international difference of responses on this topic was found significant at the level of .001 with a strong effect (Table 4.36).

Table 4.36 Number of children per household

PQ #69		1	2	3	4	≥5	Total	X^2	df	p	φ
CN	N	254	69	6	6	1	336	222.95	4	.000	.628
	%	75.6%	20.5%	1.8%	1.8%	0.3%	100%				
EN	N	40	78	54	27	30	229				
	%	17.5%	34.1%	23.6%	11.8%	13.1%	100%				
Total	N	294	147	60	33	31	565				
	%	52.0%	26.0%	10.6%	5.8%	5.5%	100%				

4.7.5 Number of adults at home

Chinese children tended to have larger families than their peers in England (Table 4.37). This suggests that even in the modernized major city like Nanjing, a large proportion of Chinese people are still following traditional family value and approximately half of the families have three to even four generations living together.

Table 4.37 Number of adults per household

PQ #70		1	2	> 2	Total	X^2	df	p	φ
CN	N	9	178	147	334	84.637	2	.000	.389
	%	2.7%	53.3%	44%	100%				
EN	N	35	166	25	226				
	%	15.5%	73.5%	11.1%	100%				
Total	N	44	344	172	560				
	%	7.9%	61.4%	30.7%	100%				

4.7.6 Number of books in the bedroom

Less Chinese pupils reported none books in their bedrooms than did their English peers; about one quarter more of them had 25 or more books in their bedrooms than did English pupils (Table 4.38). The international difference is statistically significant at the level of .001 with a modest to moderate effect.

Table 4.38 Number of books in children's bedrooms

PQ #71		*None*	*1-10*	*11-24*	*≥25*	*Total*	*X^2*	*df*	*p*	*φ*
CN	N	5	14	33	279	331	50.528	3	.000	.301
	%	1.5%	4.2%	10.0%	84.3%	100%				
EN	N	11	32	54	131	228				
	%	4.8%	14%	23.7%	57.5%	100%				
Total	N	16	46	87	410	559				
	%	2.9%	8.2%	15.6%	73.3%	100%				

Table 4.39 Pupil family background as perceived by the teacher

#	*Nation*	*1*	*2*	*3*	*4*	*5*	*X^2*	*df*	*p*	*φ*
36a	CN	6	0	1	0	1	2.813	2	.245	-
	EN	10	0	0	0	0				
36b	CN	0	1	2	0	5	7.634	3	.054	-
	EN	0	0	2	4	1				
36c *ΔΔ	CN	0	3	0	1	4	10.179	3	.017	.824
	EN	5	2	0	0	0				

Notes
1 = less than 10%
2 = [10%, 20%)
3 = [20%, 40%)
4 = [40%, 60%)
5 = 60% or more
* = Significant at the .05 level
ΔΔ = Very strong effect ($\varphi \geq .8$).

4.7.7 Home language & parental education estimated by teachers

Three sub-questions in question #36 asked teachers to estimate the proportion of pupils whose first language was different from the teaching language or dialect (36a), whose parent(s) had completed secondary or higher education (36b), and whose parent(s) had completed Bachelor degree or higher (36c).

It is clear in Table 4.39 that the TQ item #36c sees a considerable difference between two countries with a very strong effect. Apparently, more Chinese parents per class had higher levels of education than the case of England.

4.7.8 Conclusion

Chinese children were more likely to have no siblings than English children who usually had about two siblings or more. In both countries, those who had no computers at home only counted less than 8 per cent. However, the English proportion of children in this circumstance tripled the Chinese figure. Moreover, 27 per cent more children in the Chinese sample reported having more than 24 books in their bedrooms than the English proportion. In addition, according to teachers, Chinese parents seemed to have higher levels of education than English parents. The overall SES data thus seem to suggest that Chinese children might come from slightly more advantaged background than their peers in England. These pupil background differences might have affected teacher behaviours as well as children's maths performance.

4.8 Chapter conclusion

In this chapter, we have looked at the evaluative results of nineteen lessons as measured against the two internationally validated instruments – OTL and ISTOF, information of teachers and pupils collected with two questionnaires, children's test results at two points in time, pooled analyses of correlations between teacher behaviours and pupil mathematics performance, and finally pupil background differences. The quantitative part of results will join the qualitative findings (to be presented next) in Chapter 6 where pupil backgrounds and other issues – such as culture, amount of homework and teacher CPD – will be discussed at a deeper level. In the following chapter, we will move on to the QUAL part of the study to see the perceptions of different roles in and beyond classrooms regarding the quality of mathematics teaching.

CHAPTER 5 - RESULTS & FINDINGS II

Chapter overview

- *Chapter introduction*
- Unstructured lesson observations (M4)
- Interviews with teachers (M5)
- Focus groups with teachers (M6)
- *Chapter conclusion*

5.1 Chapter introduction

This chapter presents a jigsaw of mental pictures of teachers, their colleagues, their foreign colleagues and me as a new researcher. Firstly, the researcher's unstructured observations tell the typical structure and key features of mathematics lessons in each country. Secondly, interviewing data are presented and discussed to show teachers' perceptions of effective mathematics lessons, their own judgment of the quality of their observed lessons, their awareness of teaching in the other country and their attitudes and openness towards change in practice. Thirdly, around two focused lessons – one from England and the other from China – spread various views on the lessons in the wake of group discussions among teachers. Following the end of the chapter comes the final chapter in which Section 6.2.2 sees the pieces of the perspective-jigsaw from this chapter joining each other.

5.2 Unstructured lesson observations (M4)

The in-depth analyses of all lessons suggest that maths teaching tends to be similar within each country rather than across countries. There are typical lesson flows and key characteristics found across classrooms in each country. This is consistent with quantitative results on teaching. Such an intra-country similarity might be partly due to the influence of the educational policies within each country. The following sub-sections will detail the researcher's interpretation/perspective of each country's maths teaching regarding its typical *lesson flow* and *characteristics*. Firstly, the English lessons will be illustrated, and then the Chinese.

5.2.1 English lessons

A typical English lesson starts with Success Criteria, is carried out through speedy first input and long repetitive re-inputs, and ends in Success Criteria.

5.2.1.1 Typical lesson flow of the English maths class

Start with Success Criteria

After a pre-lesson practice of basic skills, such as times tables, English teachers typically start the lesson formally with a slide showing steps of *Success Criteria* after which they all name their lesson objectives. Teachers tend to be very open to share all lesson targets – big and small – with the children at the very beginning (see e.g. Figure 5.1).

Speedy first input

Children are sitting on the carpet in front of the IWB while the teacher teaches the whole class step by step the lesson objectives. At the beginning, the teacher is interacting with the whole class, but this does not last long. The teacher uses this process to screen children by ability judging by the quality of their responses or work on their mini WBs, and then from time to time the teacher sends off some of the pupils to the TA for extra support or back to their seats for the level of worksheets (often coloured according to the degree of difficulty) that the teacher thinks suit them. Those remaining on the carpet are able to progress to the next more advanced level of learning, and the ones that are last sent away are what teachers call the high-ability children. A typical maths lesson in England lasts about 50 minutes, but this selective input process only takes about 10 minutes.

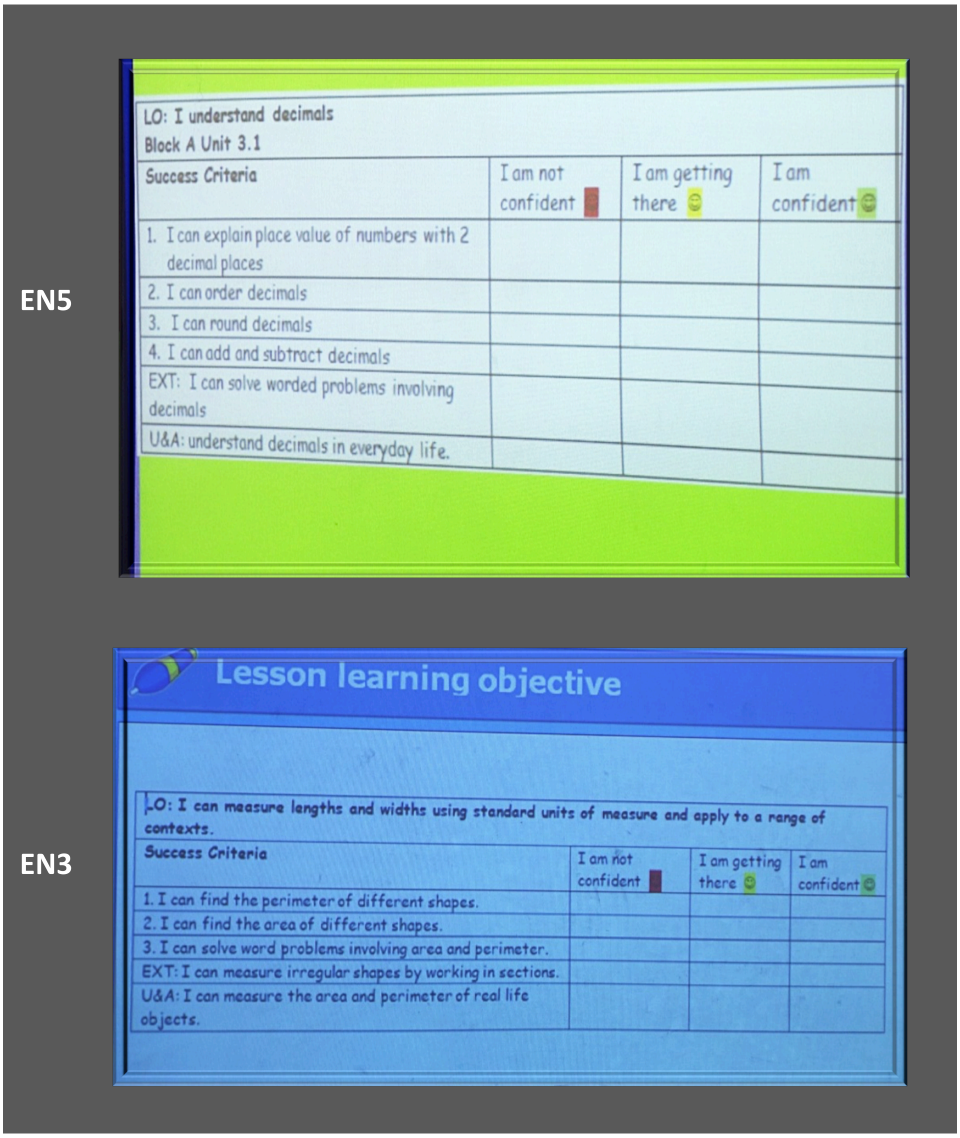


Figure 5.1 Success Criteria in English classrooms (two examples)

Long repetitive re-inputs

By the time the high-ability children finally leave the carpet for seat work, those low-ability children who have firstly been sent away are already back for extra input because they cannot cope with the tasks they have been given. Now, the cycle of "re-teaching – re-sending away – returning to the carpet – re-teaching" begins. Children at all levels keep coming back to the carpet for more support or for a new input to progress to a next (more advanced) step on the Success Criteria. Partial-class interaction all happens at this stage. From the beginning of the lesson, the teacher's attention is gradually zooming from the whole class on to individuals, and those later on fall out of the teacher's attention are supposed to be either doing independent work or self-assessing/marking their work with the pre-assigned answer sheet(s).

End with Success Criteria

During the last several minutes of a lesson, the teacher asks all pupils to gather on the carpet and reflect upon the Success Criteria steps; children then make their ways back to their seats, each self-mark their Success Criteria and stick it onto their workbooks. They are asked to bear in mind what they have achieved today and what they aim to achieve tomorrow. The lesson typically ends here.

5.2.1.2 Key characteristics of the English maths lessons

With the typical lesson flow of a maths lesson in England in mind, we will now look at the eight typical features of an English maths lesson sequentially:

- *Differentiated teaching content*
- *Learning targets unfolded at one shot*
- *Direct transmission of knowledge*
- *Repetitive inputs*
- *Weak foundation and lack of readiness*
- *Lack of accuracy in teaching*
- *Frequent physical movement of pupils*
- *Support available for special educational needs*

Differentiated teaching content

English teachers spent more time in differentiating teaching and allocated more time for children working on ability-based worksheets. Though most schools are set by ability across the same age group, in-class grouping is also commonly practised in every classroom observed. Findings from unstructured observations coincide with results of teacher questionnaire item #38.

Learning targets unfolded at one shot

English teachers show directly what are expected to learn at the very beginning of each lesson. The embedded idea is to set a destination and then get there with the destination in mind. This sounds nice and seems transparent, but the teacher might also lose the chance to nurture curiosity among pupils and make good use of such curiosity to deliver every chunk of knowledge at a right moment.

Direct transmission of knowledge

A typical approach to developing cognition in English lessons is that the teacher demonstrates or states the standard way of solving a specific type of problems, and then pupils follow the way and practise the exact procedure during independent work. This approach eventually falls into the category of direct knowledge transmission which leaves considerably little chance to a gradual development of children's cognition in a coherent manner.

Repetitive input

Because of ineffective first input of each knowledge point, individual pupils keep coming back to the carpet for extra support, hence repetitive inputs from the teacher. If the first input did not work, the teacher should have asked her/himself what else she/he could do to teach the same content more effectively before her/his first attempt to re-teach.

Weak foundation and lack of readiness

Across classrooms, English pupils show certain knowledge gaps, such as times tables, place value, subtraction and addition, which should have been built solidly enough before the observed lesson. Such gaps from time to time prevent them from advancing to higher-level knowledge, which have not however drawn much attention from the teacher, let alone motivate the teacher to stop the next new input and start to consolidate the old knowledge. When there are pupils coming back for re-input, the teacher could have already thought of two possible reasons: (1) there is something wrong with the previous teaching method; (2) the children have knowledge holes that prevent them from moving on to more advanced knowledge. The teacher should accurately predict the possible reasons that prevent children from understanding and then take action to solve the problem immediately. For example in Teacher EN6's lesson on factions, most pupils are not familiar with times tables – an essential foundation for learning fractions. Take the following interaction between the teacher and three children (anonymous) in the class as an example. To find 3/4 of 16 (Figure 5.2), one needs to be proficient with 3 and 4 times tables which happened to be the pupils' weak point:

Teacher: How many sections, Andy?
Andy: 16.
Teacher: 16 sections. OK. (writing down 16 on IWB) If I want to find 3 quarters, I need to find one quarter (writing down 1/4 of 16). How many fours in 16? (looking at the boy next to her and waiting)
(…*15 seconds* later…)
Andy: 4.
Teacher: Good boy. 4. (writing down 4) How many quarters do I need? (looking at Amy)
Amy: 3.
Teacher: So all I need to do… I need 3 lots of the answer. (pointing at 4) I've got my answer, and times it by …?
Amy: 3.
Teacher: Great. (writing down × 3 =) So, 4 times 3 is … ? (posing her fingers in a counting gesture and looking at Amy)
(…*27 seconds* later…)
4 times 3 (again with fingers in the gesture of counting), 3, 6, 9, … ?
(looking to Amy)
Amy: 12.
Teacher: 12. (writing down 12 behind the "=")

Probably because of heavy reliance on the calculator from a very young age, English children show considerably weak calculation skills and number sense across classrooms. Teachers often suggest them to count fingers, which results in many children doing so in the process of completing tasks (e.g. see Figure 5.3).

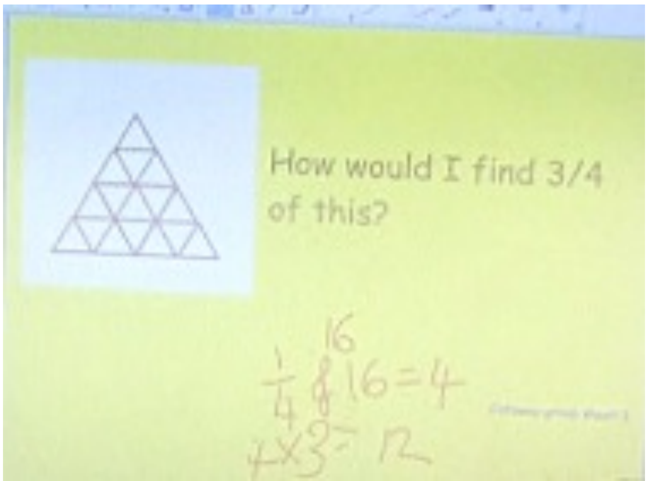


Figure 5.2 Task example from Lesson EN6

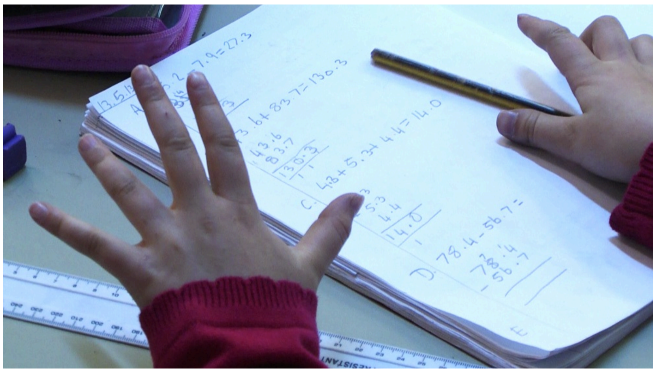


Figure 5.3 English children counting fingers for calculation

Lack of accuracy in teaching

Across classrooms teachers tend to write down inaccurate statements or answers, and their handwriting also tends to be informal.

For example, in Teacher EN6's lesson, to represent part of a whole, the teacher sometimes wrote it as 1/3 of 18 = 16, which still made sense though it should be formally written as 1/3 × 18 = 6. But, sometimes the teacher simply omitted "of n", for example, writing 1/3 × 12 = 4 as 1/3 = 4 (Figure 5.4).

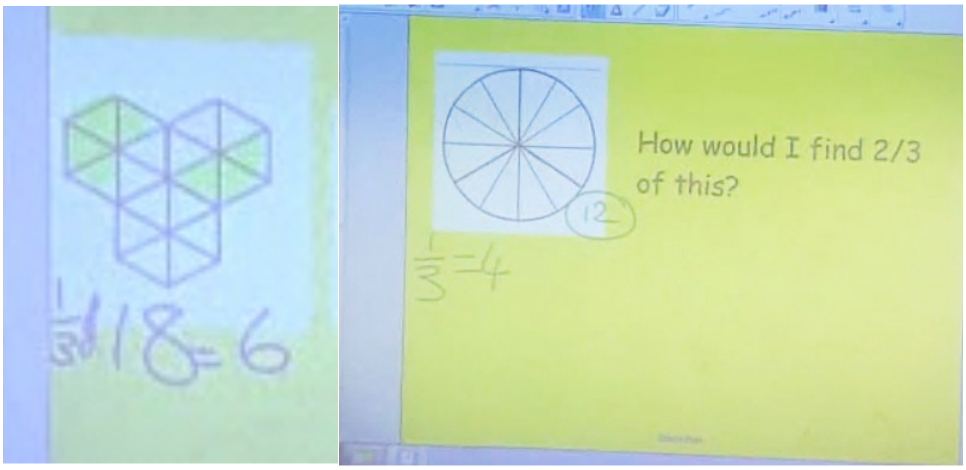


Figure 5.4 Inaccurate expressions on the board (EN6)

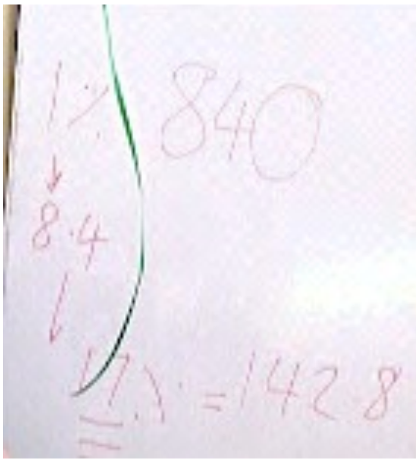


Figure 5.5 Wrong percentage sign and wrong equation (EN8)

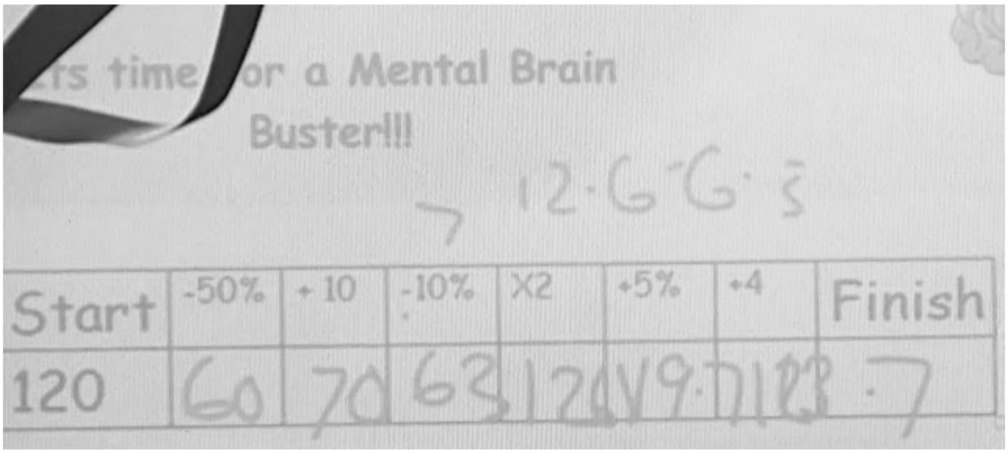
Similarly, when Teacher EN8 was trying to demonstrate 17% of 840 was 142.8, she made two mistakes: 1) an inverse percentage sign after the number 17 and 2) an incomplete expression which should be written as 17% × 840 = 142.8 rather than 17% = 142.8 (Figure 5.5).

Such inaccurate writing on the one hand may cause confusions among pupils in the specific lesson, and on the other may also set up wrong examples for children to copy. If one day the teacher finally spotted the inaccurate or wrong expression in children's work, she/he would have no idea when, where and how the misconception had been formed.

There are also complete mistakes. For example, at the end of the lesson EN8, the teacher was explaining the answer to the final task. Everything went fine until it was time for the 6th column which demanded an increase of 5% from the previous answer 126 (Figure 5.6). The teacher took "+5%" as "–5%", and none of the pupils found it wrong or, if they did, none of them pointed it out. Of course, the final result was absolutely wrong. If she had prepared the answers beforehand, she would have realised the difference, but apparently she had not.

The bottom part of Figure 5.6 shows a pupil's work – randomly captured by the camera. It was wrong from the fourth column (i.e. –10%) onwards. Apparently, the pupil was not completely clear about the rule of the game, which had not been spotted by the teacher and probably would never be. When she wrote the final "answer" in the last cell, the teacher announced that the lesson would stop here and "Table 2" would be dealt with some time after the lesson.

Without clear handwriting and correct demonstration, any intended content would not be successfully understood and grasped by pupils. The results might be even worse – the lesson time had been wasted, more confusion had been embedded and new learning obstacles had formed.



Unidentified errors in a pupil's work in Class EN8

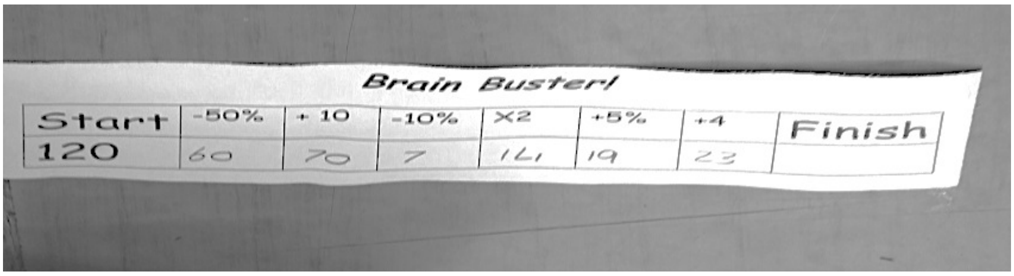


Figure 5.6 Wrong answers & poor handwritings on the board (EN8)

Frequent physical movement of pupils

Activities happening in the English lessons often involve the change of pupils' seating/location in the classroom. In every classroom, pupils move from their seats to the front carpet at the beginning and end of the lesson for whole class interaction and during the middle of the lesson for multiple times mainly to receive re-inputs from the teacher either individually or in a group. This seems similar to what American teachers called "controlled chaos" (Cai et al., 2014, p. 274).

Support available for special educational needs

Apart from the classes EN3, EN4 and EN10, all other English classes had at least one teacher assistant (TA) present to offer extra support to children lagging behind or with special educational needs. In Teacher EN5's class, there were two to three TAs. This was different from the case of the Chinese classroom where only one adult – the teacher – was available for all.

Having seen the flow and characteristics of maths lessons in England, we will next look at the Chinese maths lessons.

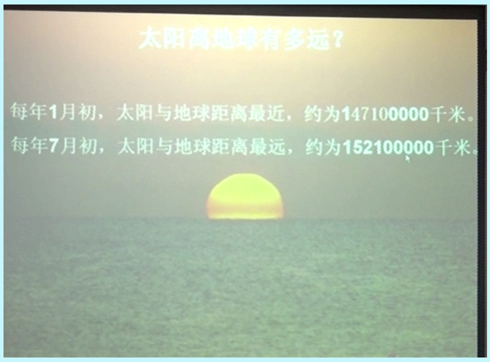
5.2.2 Chinese lessons

A typical Chinese maths lesson starts from and ends in real-life situations, and in between is a systematic problem solving process in which the teacher plays a less active role and the pupils need to think and discuss actively the solutions to a sequence of tasks. And teachers seem to take the sequence seriously.

5.2.2.1 Typical lesson flow of the Chinese maths class

Start from real life

Chinese teachers tend not to tell what is in the bag at the start of a lesson (Figure 5.7). They all started from real-life issues or stories that were relevant to both pupils' life and the lesson content, posed open-ended questions and then waited for the "right moment" to come when they would point out the title of the lesson. The intention was obvious – to draw pupils' attention without being too eager to load the burden of the whole-lesson's objectives onto pupils' minds at this stage.



Translation

Counting numbers with stones

Translation

What is the distance between the Sun and the Earth?

At the beginning of every January, the distance is the shortest – approximately 147100000 km.

At the beginning of every July, the distance is the longest – approximately 152100000 km.

Figure 5.7 Real-life issues as 'starters' in Chinese maths lessons

For instance, to teach the lesson on place value, Teacher CN4 started his lesson from the way ancient people counted sheep, using stone substitutes and the idea of base ten. In a consolidating lesson on numbers from ten thousands to hundred billions, Teacher CN9 started with a riddle about the Sun and a discussion about the distance between the Sun and the Earth. Teacher CN8, to start her lesson on

Using Strategies to Solve Word Problems, asked pupils to reflect on and talk about the Aesop's fable, The Crow and the Pitcher, which led to the key word of the lesson – "strategy".

Task-oriented process with systematic reasoning and problem solving

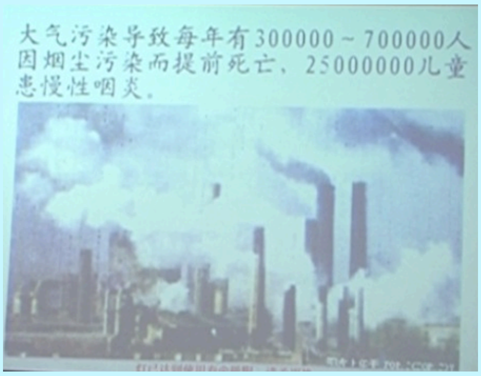
The main part of a Chinese lesson is a gradually unfolding process in which the teacher promotes children to advance step by step through the lesson content. At the heart of the content are two to three exemplary problems that underpin the intended subject knowledge. The class only deal with one problem at a time, from simple to complex.

Each exemplary task generally involves five steps: (1) the teacher poses the problem and pupils discuss it with their desk-mates quickly; (2) the teacher interacts with the whole class through intensive Q & A; (3) one or multiple solutions are generated by the whole class; (4) children try the solution(s) with a similar task and the teacher quickly circulates through the class to check everyone's work and collect representative work examples; (5) the teacher shows pupils' work examples – both correct and wrong – to the whole class via a projector and initiates whole class discussion regarding the answers.

Steps (2) and (5) usually take longer time than other steps, which allows pupils to thoroughly practise reasoning and metacognitive thinking on the specific type of tasks, hence consolidating the newly obtained knowledge in time. Step (5) is particularly helpful in enriching pupils' learning experience with typical correct and wrong solutions carefully chosen by the teacher during seat work. Step (5) also involves lots of reasoning on why certain solutions are right or wrong and how an alternative or wrong solution has been produced. Every single task thus demands sufficient time and thorough consolidation, and once it is done, there is no playback.

Translation
In 2009, the total of car emissions was 51430000 tons.

(Class CN4)



Translation
Air pollution is the cause of 300000 to 700000 deaths and chronic pharyngitis of 25000000 children each year.

(Class CN5)

Figure 5.8 Real-life situation as endings in maths lessons in China

End in real life

Chinese lessons are all linked back to real-life situations to allow pupils to think about why and how the mathematical knowledge may relate to the real world and/or help solve problems in real life. For example, to end lessons on higher-level numbers (ten thousands to ten millions), Teachers CN4 and CN5 showed pictures about air pollution (Figure 5.8) to the class and asked them to share their thoughts.

5.2.2.2 Key characteristics of the Chinese maths lessons

Chinese maths lessons carry eight main features which are as follows and will be introduced sequentially afterwards:

- *Pupils as active thinkers*
- *Wide existence of metacognition*
- *Solid foundation and absolute readiness*
- *Pupil classwork as teaching resources*
- *Optimal timing of teachers' writing on the board*
- *Teacher-guided and pupil-centred whole-class discussion*
- *Less physical movement of pupils*
- *Big class size and lack of special educational support*

Pupils as active thinkers

Chinese teachers are more active in posing questions so as to generate solutions among pupils but less active in providing solutions. The duty of looking for an answer or solution to a given problem is on the pupils' shoulders. The degree of a Chinese teacher's input and support gradually decrease as she/he guides the class through the lesson. For instance, in Teacher CN4's lesson, the exemplary task involved the collaboration between the teacher and the pupils; the remaining three tasks saw a decrease of the teacher's input and an increase of pupils' independent effort; for the last task, the pupils worked in an absolutely independent fashion and completed it over a very short period of time – 33 seconds, all done!

In Chinese classrooms, the majority of the lesson time is spent on whole-class reasoning and discussion about questions as to how to solve the exemplary tasks, whether there are alternative methods, why certain emerging answers are correct or wrong and which step goes wrong if an answer/solution is incorrect. Pupils learn from each other's experiences throughout the lesson, and the teacher always sees the whole picture of pupils' knowledge development and their readiness for the next bit of teaching and learning. Moreover, in Chinese lessons, there is constantly indirect reasoning about why mathematics should be learnt, how mathematics is connected with the wider world and pupils' life and how knowledge obtained in the classroom might be applied in the real world. Children are motivated to learn actively all the time.

Wide existence of metacognition

One of the typical features in Chinese classrooms is the existence of lots of metacognitive discussion promoted by the teacher and carried out by the pupils particularly before and after each task. Before tackling problems, pupils are much concerned with issues as to what they must bear in mind when solving problems and why. This is often generated through intensive questioning and reasoning in the whole class. Afterwards, still motivated by the teacher questioning, pupils think and talk about issues as to what the appropriate methods and correct results are, why certain mistakes have been made by themselves or other pupils and what measures should be taken to avoid similar mistakes in the future.

Solid foundation and absolute readiness

Across all Chinese classrooms, children show their solid foundation and absolute readiness for the new content to come in the observed lessons. Chinese teachers seem to be more aware and confident of their pupils' level of knowledge and thus make more accurate 'prescriptions' for their capacity of learning in this new lesson. If a teaching plan is a hypothesis, then Chinese teachers are good at making accurate hypotheses.

Pupil classwork as teaching resources

The use of pupil classwork in the whole class discussion helps enrich pupils' learning experiences through observing and thinking about others' learning experiences. For example, in his lesson on *Knowing Numbers at the Wans Level* [5], Teacher CN5 showed the exemplary representations of 195 0000 and 200 0000 [6]

---

[5] In Chinese numerical system, the wans level (万级) consists of four digits: 万位 wans (ten thousands), 十万位 ten wans (hundred thousands), 百万位 hundred wans (millions) and 千万位 thousand wans (ten millions). See the English-Chinese corresponding column names below:

English column headings	*Ten millions*	*Millions*	*Hundred thousands*	*Ten thousands*
Chinese column headings	Thousand wans 千万	Hundred wans 百万	Ten wans 十万	wans 万

[6] The Chinese numerical system separates numbers by four digits rather than three. For example, 1,000,000 is read out by separating the number like 100 0000.

operated by pupils on the abacus. Figure 5.9 also shows snapshots about two teachers presenting pupil work to the class via the projector and two pupils explaining or demonstrating their work to the class.

Numbers 195 0000 and 200 0000 shown in Class CN5

Class CN1

Class CN8

Class CN2

Class CN5

Figure 5.9 The use of pupil work in Chinese classrooms

Optimal timing of teachers' writing on the board

Chinese teachers all use both digital and physical (traditional) presentations as a way of representing mathematical content in their lessons. Of either approach, the typical feature is that they time 'accurately' (according to their own plans) when to write what on the board. The contents on the board by the end of a lesson would always be but not limited to the title of the lesson and the structure / key points of the lesson. Essential words will only be written on the board at the point when the teacher feels key ideas have emerged naturally during the whole-class discussion.

Teacher-guided and pupil-centred whole-class discussion

In a Chinese maths class, the teacher's role is more like a guide who ensures that the teaching and learning is heading in the right direction and who at the same time leaves the main job of mathematical reasoning and problem solving to pupils themselves. It was through whole-class discussion that the guidance of the teacher and the central role of pupils simultaneously come true. Chinese teachers are always encouraging pupils to open their mind and come up with multiple solutions to one problem. In this process, children have opportunities to think and tackle a given problem from different angles. The whole interactive process thus takes their thinking beyond the surface level of the problem and deepens their understanding of the intended mathematical knowledge.

Less physical movement of pupils

Contrary to the case of England, pupils rarely move from one place to another in any Chinese classrooms. Order seems particularly important in Chinese classrooms, and the collective effort between the teacher and pupils in keeping the order is an observable habit. There are several occasions in each Chinese classroom when individual pupils are asked to come forward to write answers/solutions on the board or manipulate instructional tools, such as the abacus, but only for the purpose of whole-class discussions.

Big class size and lack of special educational support

The Chinese classroom has a bigger class size than the English classroom. Apart from Class CN-7 which had 28 pupils at the time, the size of all other eight Chinese classes ranged from 35 to 43. Although in China disabled children all attend special education schools, there might be some pupils attending state schools who at times also need extra support to meet their needs. The large class size seemed working fine in all participating classes in the study, but the focus of the only teacher on the majority of children in one class meant several low-ability children might possibly easily be ignored in the teaching and learning process. There was clearly the lack of special educational support in place ideally from a teacher assistant across Chinese classrooms.

Now that the researcher's unstructured observations of two countries' maths lessons have been presented, the following section will provide a picture of teachers' individual views on their teaching in general and observed lessons in specific.

5.3 Interviews with teachers (M5)

Teachers' opinions are clustered primarily under the interviewing questions within each nation and subsequently synthesised across nations wherever necessary.

5.3.1 Teacher beliefs

This section will sequentially present national pictures of (1) teacher beliefs of what counts as an effective maths lesson, (2) teacher strategies of organising a maths lesson, and (3) teacher strategies of differentiating teaching.

5.3.1.1 Description of an effective maths lesson (itvQ1)

English views on an effective maths lesson

In English teachers' mind, an effective maths lesson should be (1) relevant to real life, (2) differentiated and (3) child-centred.

Relevant to real life

Teacher EN1 thought that in an effective maths lesson everyone should be able to make progress and apply what she/he has learned. He further put it in this way:

> It's not just based in the classroom. They can think outside. And it's the application of it really. For example, decimals, they can apply it to money. … It's like connecting things up to apply them to the real-life situations.
>
> (Teacher EN1, 26 February 2013)

Teacher EN4 also thought a lesson should ideally be the one that had a real-life meaning. Likewise, **Teacher EN10** said:

> Maths should be as inspiring and as interesting as possible. There are so many jobs, you know, the children will go on to do. They need maths for them. I think if they are not interested in maths, they are not gonna have access to those careers.
>
> (Teacher EN10, 8 May 2013)

Differentiated

In **Teacher EN2**'s view, an effective lesson should be appropriate to what the children already knew and be challenging enough to help them learn something new. The lesson should also be based in assessment of what children's targets were and the levels of curriculum they were at.

Similarly, both **Teacher EN4** and **Teacher EN8** also thought that an effective maths lesson should enable all children learn at the level that was right for them, make progress and think about what their next steps of learning were.

Teacher EN10 also pointed out that activities should be differentiated to suit various-ability children's needs. Likewise, **Teacher EN7** thought the necessity of knowing that "not every single child is going to learn in the same way". Similarly, **Teacher EN9** saw differentiation as the soul of an effective lesson.

Teacher EN6 also regarded an effective maths lesson as the one appropriate for their levels and involving self-assessment by the children themselves. Nonetheless, she also emphasised the importance of the on-going assessment by the teacher, as this would help make the differentiation clear.

It seems that English teachers universally appreciate the need of teaching with differentiation and the necessity of assessment-informed differentiation. This is consistent with what have been observed in their lessons.

Child-centred

Teacher EN3 expressed that an effective lesson should offer lots of opportunities for children to learn and do not involve too much teacher talk. In his view, the lesson should have a good pace, more interaction among pupils, clear objectives at the beginning, and self-assessment at the end. **Teacher EN2** thought behaviour was not an issue and in an effective lesson:

> the child should be involved in setting up targets, deciding where they start working or they are going to work on, which motivates the children, because they are not told to do something. There's no ceiling to the children, and they are free to move.
>
> (Teacher EN2, 26 February 2013)

Teacher EN5 described an effective lesson as the one that engaged children and moved them along in their learning. **Teacher EN10** thought that in an effective maths lesson all children should be able to make progress, whether they were of high or low ability. He stressed that such a lesson should have input for every child in the class.

Chinese views on an effective maths lesson

In Chinese teachers' mind, an effective maths lesson should (1) be aimed for and well accepted by pupils, (2) have real-life and long-term meaning, (3) be well planned and (4) have the lesson goals achieved.

Aimed for and well accepted by pupils

Teacher CN2 considered the golden rule of assessing the effectiveness of a lesson was to see how pupils reacted in the class. **Teacher CN3** thought an effective lesson should be delivered in a specific method and easier for pupils to understand immediately. She emphasised that such a lesson should also give opportunities to children for them to showcase their work, be their work correct or wrong. **Teacher CN10** said that children should be enabled to learn the knowledge *solidly* in a happy and relaxing mood. **Teacher CN1** thought an effective maths lesson should take both the environment and the specific learners into consideration. Ideally there should be a very high level of active learning. She also thought the effective teaching must consider the differences among pupils.

Teacher CN8 said such a lesson should "shine with the glory of humanity" and put children at the heart of it.

Real-life and long-term meaning

Teacher CN4 thought an effective lesson should be helpful to the children throughout their life and add up to their ability in meeting challenges in their future work and life. **Teacher CN3** said, before the lesson, an effective teacher should have clear answers to:

(1) What is the context of the knowledge?
(2) What are the connections between knowledge and children's life?
(Teacher CN3, 11 December 2012)

Teacher CN1 expressed that the short-term goal was to get the pupils to grasp the methods of learning, and the long-term goal was to build on the short ones to help them form good learning habits. Thus, the children would still be able to apply the attained methods and habits in future learning and continue progressing on their own. Likewise, **Teacher CN5** thought it should enable pupils to synthesise knowledge and apply what was learnt in their life. **Teacher CN3** also thought children should be able to apply the knowledge they obtained.

Well planned

Teacher CN5 thought to make a lesson effective the teacher had to prepare the lesson well but be flexible during the lesson. **Teacher CN3** pointed out, to make a lesson effective, a teacher should bear in mind clearly:

(1) What are the key points and the difficult points of the lesson?

(2) What approaches should be taken to address the lesson content?
(Teacher CN3, 11 December 2012)

Lesson goals achieved

Teacher CN6 put it simple, "If the intended results were achieved, then the lesson was effective." Similarly, **Teacher CN7** thought in an effective lesson the teacher should have completed the lesson objectives, pupils should have grasped what was taught and their ability should have been well developed. **Teacher CN8** thought an effective lesson should equip pupils with the strategy of solving given problems. **Teacher CN9** thought, before an effective lesson, the teacher should have listed clearly the lesson objectives and, after the lesson, the teacher should be able to see that pupils had fully achieved all those objectives.

Having heard teachers' answers to the first interview question – what counts as an effective maths lesson, we now move on to the following subsection which shows their daily strategies for organising maths lessons.

5.3.1.2 Strategies for organising and managing lessons (itvQ2)

English strategies for organising and managing lessons

English teachers all came up with one and only approach to organising and managing maths lessons which is assessment-aided differentiation.

One theme: Assessment-aided differentiation

Teacher EN1 thought the ability-based set was helpful for the TA to come over and help a group if there were any maths misconceptions. He anticipated that it would be more than likely that someone of the similar ability had got that misconception as well, so then the TA could address the whole table. He stressed the importance of independent learning in maths, for example, providing pupils with an answer sheet so they could self-mark their work.

Teacher EN2 generally started a new unit with a quiz to pre-assess where children were so that they could decide which step was appropriate to start on the Success Criteria. She would then normally focus on a group of pupils and assess whether they were aiming too high. In the end of a lesson, she would use a mini-challenge to assess how far children had made during the session.

Continuing with his emphasis on differentiation, **Teacher EN3** further clarified that his first strategy was to try to differentiate his class (the top maths set of the year group) in three ways: either A, B and C (or lower, core and upper activities). He often chose to be with the lower activity, which he said was due to the lack of a TA in his class. The second strategy of his was to do what he called “prior learning” at the start to see what children could and could not do. This helped him to decide who could do what.

From the same school as **Teacher EN3**, **Teacher EN4** also emphasised the importance of knowing the learning objectives and breaking them down to steps of learning in the Success Criteria. To know who could do what, she would generally do a prior activity to group children according to their levels. Also not having a TA, she expressed the difficulty of meeting every child's needs at different levels as she planned. Her follow-up strategy was:

> what I do is I do a record of where they are at the end of each lesson so that I know the next day what learning they need or what group they need to be in.
>
> (Teacher EN4, 9 May 2013)

Teacher EN6 usually started with a "mental starter" to get the pupils thinking and then moved on to the main activity with separate inputs serving for children of "a wide-range of ability". She generally assessed children's WBs on the carpet and then sent some away to independent work or brought them back for the next step of learning.

Teacher EN7's strategy was to assess children's levels at the beginning of each week so that they could be grouped by ability:

> If the children get every single one right, they'll be on the top table; if some get a few of them right, they'll be on the middle table; if none of them, they'll be on the bottom table.
>
> (Teacher EN7, 1 May 2013)

She would then plan a week of lessons to move the pupils' learning on. Having said that, she pointed out that her plan was flexible and everything depended on how pupils got on and how she taught them. "So I always change," she said.

Teacher EN8's strategies were to assess children at the beginning of the lesson to have an idea "where children already are". She would then differentiate her teaching accordingly. **Teacher EN9** also saw differentiation as the core of her strategy, though she pointed out that it was important to be flexible at the same time.

Teacher EN10's main strategy was differentiation through the "cutaway style" as called by teachers at the school. With this method, the teacher grouped children by assessing and sending some of them away for independent work at each step at the start of a lesson. This often happened on the carpet where children's work on their mini WBs would be swiftly viewed and assessed by the teacher.

Chinese strategies for organising and managing lessons

Chinese teachers offered a wider range of strategies than did their English colleagues. These strategies include (1) paired or group discussion, (2) the use of warm reminder, jokes or riddles, (3) matching the characteristics of each lesson, (4) promoting peer instruction among pupils and (5) being well prepared and flexible.

Paired or group discussion

Teacher CN2 usually set paired discussion as a way of organising lesson activities. **Teacher CN3** used different ways of organising activities depending on the purpose and types of lessons. Usually it would be paired discussion, whereas groups of four would be arranged to collaborate on tasks in a practical lesson. In terms of grouping (*not* by ability), **Teacher CN4** would ask pupils to form a group of four on explorative tasks, but generally arrange paired collaboration. **Teacher CN5** said he generally broke tasks into small chunks for pupils to work on and, if there seemed to be any difficulty, he would ask them to discuss in pairs. **Teacher CN6**'s strategy in organising activities was also through paired discussion and/or collaboration.

The use of warm reminder, jokes or riddles

In terms of discipline, **Teacher CN3** said she generally gave a warm reminder to an off-task pupil, *but* she would reflect upon her own teaching if there were *lots* of off-task pupils in the class. **Teacher CN4**'s strategy was to make the lesson more interesting and less dull, through telling jokes and stories and using real-life situations. If there were bad behaviours, he would generally send information through eye contact or remind the pupil in a humorous tone. **Teacher CN9** would generally use a game to boost pupils' interest, for example, asking them to guess a maths riddle, so that they would wonder why the teacher did so and what might be happening in the lesson.

Matching the characteristic of each lesson

Teacher CN7 said she used different strategies in different lessons. In a calculation-oriented lesson, she would mainly ask pupils to do practice; in a lesson at the end of a unit, she would set more games and activities; in a practical lesson, she would ask pupils to interact with their desk-mates or in a group of four on a more complex task.

Promoting peer instruction among pupils

During the Q & A process, if a pupil gave a wrong answer, **Teacher CN7** would paraphrase her question and try to guide the pupil to get the correct answer independently; if it still did not work, she would ask others to help the pupil, rather than explain the correct answer directly herself. She insisted that pupils, instead of the teacher, should find the answer. For the development of those pupils who did not grasp the knowledge after thorough instruction, **Teacher CN9** would show their work to the class via the projector during the whole-class discussion and ask other pupils to help them identify reasons for their mistakes and offer solutions.

Being well prepared and flexible

Teacher CN8 generally made a very detailed plan and was able to make accurate predictions of the lesson, therefore coping freely what emerged in the lesson. Having said that, she thought the strategy should depend on the characteristics of each lesson. **Teacher CN9** usually took time to design the questions in a clever way so that every pupil could come up with a few ideas to share with the class in the lesson.

Overall, English teachers tended to have a similar strategy in organising and managing their maths lessons, whereas Chinese teachers had more ideas and more detailed purposes in doing so. The next subsection tells their strategies for differentiation in maths lessons.

5.3.1.3 Strategies for differentiation in maths lessons (itvQ3)

English strategies for differentiation

English teachers talked about four different strategies for differentiation including (1) self-differentiation, (2) pre-assessment, (3) support from TAs and (4) different abilities different tasks.

Self-differentiation

Teacher EN2 explained that, since the school did not have groups of children, the steps in Success Criteria provided the opportunity for children to move through according to their self-judgement. This coincided with her colleague Teacher EN1's words – "lots of self-differentiation" in the class.

Pre-assessment

Teachers EN3 and **EN4** both took "a prior learning activity" they called as the first step of differentiation in a lesson. **Teacher EN3** said this activity helped him to see what pupils could do from the previous year group. Then, he would check pupils' answers by quickly looking at their WBs, so that he could immediately see which group needed more challenge and which group needed to "be brought in support". After this, he would decide which group he should be working with. **Teacher EN4** also took the prior learning activity as a key step towards differentiation. She said:

> I sort of differentiate from there (i.e. the prior learning activity) – from that point. So whatever they need – if they need larger numbers or need to do a completely different method or … that sort of employ my differentiation. In today's lesson, I have three differentiated tasks, but they weren't secure enough to leave – some of them – so I have to keep them for a bit longer than I wanted to.
>
> (Teacher EN4, 9 May 2013)

Teacher EN6 mentioned a maths app that helped her break down "each strand of maths into the levels" and gave her a rough idea of children at different stages. At the beginning of a lesson, she would quickly check children's WBs or reflect on their pervious work and decide where they needed to go.

Teacher EN7 continued with her weekly strategy in managing the class to detail her differentiation approach. She reemphasised that the beginning activity set at the start of a week helped her decide which group the children fell in. Then, the activity on the carpet in each lesson further helped her differentiate the kind of support she gave to different children. Having said that, she still found it "quite hard to differentiate".

Support from TAs

Teacher EN3 showed a sign of helplessness for not having "the luxury" of a TA. He said, if he did, he would ask the TA to work with another group – "a target group". His school was short of TAs, and at that point the school's TA-child ratio was one to 1/3 of the number of children. Conversely, it was not mentioned by him but was observable that his neighbouring class - **Teacher EN4**'s – had usually two and sometimes three TAs. Like Teacher EN3, Teacher EN5 also did not have a TA which she thought would make a huge difference if she had got one.

Teacher EN5 continued with her answer to the previous question about how well a daily plan worked for her than a plan by "block". She could get more accurate ideas of what children were capable of doing, what they could not and what they needed to move on to. However, no matter how well she prepared, there was always a group who could not:

> just remember how to do it sort of thing. So they'll work independently or in pairs, depending on how confident they are.
>
> (Teacher EN5, 14 May 2013)

When being asked whether she still set by ability within this already bottom-set class, she explained:

> It varies. Mine is very fluid grouping, because it all depends on how they performed the day before as to which group I put them in the next day.
>
> (Teacher EN5, 14 May 2013)

Teaching a lower set in another school, Teacher EN9 also expressed her reliance on her TA to help meet different needs of children.

Different abilities different tasks

Teacher EN8 found within her middle-set class "still a big gap" among pupils ranging from low level 3 to high level 5. Her differentiation was mainly between the sheets that children were doing and the amount of support that she gave to them:

> Do they need a quick chat with me? And move away? If they don't understand it, is that to break it down? Let's do an example on the board, and try one on your book. Still don't understand [it]. Let's do another on the board. Try one on your book. See how they go with it. I mean, there's a lot of time – some of the children on the carpet for a long time – but they're still doing their work.
>
> (Teacher EN8, 1 May 2013)

Teacher EN9's strategy was to differentiate her expectations to different ability children, expecting higher-ability ones to use "a variety of skills and strategies" and those with lower ability to use probably only one strategy. **Teacher EN10** would also give different objectives – easier or more difficult examples – to the pupils in the top set, so that "they were all doing the same work but at different levels" that suited them.

Chinese strategies for differentiation

Chinese teachers seem to have quite different strategies for differentiation from English teachers'. Their strategies are (1) aiming for the majority, (2) questioning techniques, (3) post-lesson support to the low ability and special awards to the high ability, (4) put the high and low abilities in pairs or groups and (5) low ability pupils to preview lesson contents.

Aiming for the majority

All Chinese teachers mentioned that their lessons were generally tailored for the majority of pupils. Some pupils already got the knowledge before being taught in the class, so the teacher would encourage these high-ability children to speak up more to set the example for mid- and low-ability pupils. For those very few low-ability pupils, extra support would be provided after class.

Questioning techniques

Chinese teachers all described similar techniques of questioning: (1) for normal questions, children would be asked randomly; (2) for simple ones, low-ability children would be asked, so that they could feel they were at least capable of *some*thing, that they were more likely to stay on task, and that their self-esteem could be uplifted; (3) for difficult questions, high-ability pupils would be asked. For example, **Teacher CN1** talked about her strategy in dealing with difficult questions. After a high-ability child correctly answered the difficult question, she would firstly ask those slightly lower than the high-ability pupils to re-explain the process of solving the problem, secondly ask the whole class to explain the solution to each other in pairs, and finally ask a low-ability child to talk about the solution again to the class. Thus, the teacher was able to consolidate the same subject content amongst a variety of learners, through asking different pupils different levels of questions and asking different pupils to answer a same question at different points in time. **Teacher CN9** also said she would first ask high-ability pupils to do a challenging task and then invite middle- and low-ability children to re-explain the solutions so that the latter two levels of children could "absorb and transform the knowledge into their own" through the re-explanation.

Post-lesson support to the low ability and special awards to the high ability

As shown in the teacher-questionnaire results, the ability range in Chinese classrooms is quite narrow which is confirmed by the tests results. For quite few lowest ability pupils, all teachers' strategy was to mark and explain their homework to them in person. For example, after the lesson, **Teacher CN9** generally asked other pupils to help the low ability, or she would help them herself if they still had some problems with their work. For those high ability pupils, if they achieved 98 or 100 per cent in any unit tests or mid-term tests, **Teacher CN3** said they would be issued a "gold" medal which guaranteed a homework-exemption opportunity. The teacher said however none of them had used the medal once this term (i.e., September 2012 to January 2013), by which she seemingly meant those pupils actually very much wanted to do homework. In some of her lessons, the high ability children might be given challenging tasks, or be authorised to read their favourite books while others were doing exercises.

Put the high and low abilities in pairs or groups

In some classes, Chinese, Maths and English teachers would discuss together as to who should sit next to whom according to their ability levels. They sought to put high- and low-ability pupils in pairs, so that the high ability could help their low-ability buddies.

Teacher CN7 said her strategy was to organise pupils' seats through grouping them by four, with each group consisting of a high-ability pupil (#1), two middle-ability ones (#2 and #3) and a low-ability (#4). She would give specific questions during the lesson – challenging questions firstly for #1 to answer, then #2 and #3 to give supplementary answers and finally for #4 to conclude the answers. Conversely, about simple questions, the teacher would ask #4 to talk firstly, #3 and #2 to follow up secondly and finally #1 to give complementary/alternative answers.

Low ability pupils to preview lesson content

Two teachers from School CN-D, **Teachers CN8** and **CN9**, both mentioned their strategy of suggesting low-ability children to preview lesson content. During the lesson, **Teacher CN9** said she would also give more opportunities for low-ability children to answer questions and respond to them with more positive feedback to keep them confident.

Summary

Section 5.3.1 focuses on three strands of teachers' beliefs: (1) a description of an effective maths lesson, (2) the strategies for organising and managing lessons and (3) the strategies for differentiation in maths lessons. Overall, English teachers tend to have a narrower range of answers for all three questions than Chinese teachers; English teachers are also more concerned with differentiation and the utilisation of assessment for differentiation than are their Chinese colleagues who focus on the majority and seek to use questioning and methods other than assessment to promote the progress of different-ability children; Chinese teachers demonstrate deeper thinking than English teachers when talking about the above second and third strands of teacher beliefs.

The following section, Section 5.3.2, shows teachers' comments on their own lessons while watching the lesson videos.

5.3.2 Teacher self-evaluations

This section contains three components regarding teachers' (1) reflection on the teaching plan, (2) self-evaluation of the implementation of the plan and (3) self-comments on the lesson and its effectiveness.

5.3.2.1 Reflecting on the teaching plan (itvQ4)

English reflection on the teaching plan

Teacher EN1 said the first thing was every pupil should write the date on their book and know what they would be doing and then he would ask them questions about their marking on their workbooks the day before the lesson. Following that was a quiz that would assess children and tell them when they should leave the carpet and work independently on certain steps (i.e. differentiated tasks). Then the teacher would work with the group that needed more support.

Teacher EN2 planned the unit – times tables – as a whole and the lesson observed was amid the unit. She was concerned that children at the School EN-A were not good at times tables. The school was in the process of improving it. This unit had been taught and learnt last year but this year it was retaught and relearnt. The teacher could see children were enjoying timing and challenging themselves and believed that they would be able to make it through overlearning.

Teacher EN3 planned to let the children practise their times tables for about 8 minutes as a "starter". Then, he would do a 10-minute prior learning activity and then a 30-minute main activity on the calculation of perimeters and areas of different shapes and "ideally" bring pupils together in the last ten minutes to allow them to self-assess their Success Criteria and stick it to their books. Children would put their books into three trays of different colours, with green meaning "I'm confident", yellow "I'm getting there" and red "I'm not confident". He thus planned to circulate the class thoroughly so as to spot every possible mistake. For example, circulating in the lesson observed, he found many pupils missed certain sides while calculating the perimeters of irregular shapes (L-shapes).

Teacher EN4 planned to start with *a multiple of ten times by a unit number* and then *a multiple of ten times by a multiple of ten*. "The class has done this before so it would be a recap," she said. She thus hoped the children would be quicker than they were previously, so she had not planned to do a prior activity to assess where they were.

Teacher EN5 originally planned to start with a times tables work and then a quick activity to practise a skill from yesterday – rounding up or down decimals to put it into money rather than just random decimals. The main part would be adding and subtracting decimals. The ending would be children self-assessing how they felt they had done, what they thought they needed to do and why they said they were just "yellow" (getting there) rather than "green" (confident).

Teacher EN6 had done a lesson four days before the observed one on the same content – fractions of a shape. In this lesson, the low-ability children were expected to work out 1/2 and 1/4 of a shape, the middle ones tried to figure out 1/5 and 1/6 of a shape and the top ones were asked to find 2/3 and 4/5 of a shape where the numerator was more than one. Four days later, the teacher planned to do a quick recap and then immediately move on to "trickier" factions of a shape where, for example, the sides or segments were different.

In the lesson prior to the observed one, **Teacher EN7** noticed that a group of children did not get to the point she wanted them to, so she planned to send this group to the TA in this lesson. She would work with the main group on the carpet about ordering decimals and factions.

Teacher EN8 aimed to teach the low- to middle-ability children to calculate 50% of a given number and high-ability ones firstly to figure out 1% as well as any percentage of a given number, secondly to take a deduction off a number and finally to solve word problems involving these calculations.

Teacher EN9 planned to teach decimals, having taught the children about fractions in the prior lesson. First thing she planned to do was to use sectioned shapes to show children that a decimal, like a fraction, was also a part of a whole. Secondly, she planned to do decimal number lines to help children recognise the sequence of decimals, such as 0.1, 0.2, and 0.3. Thirdly, the focus would be on place value of decimals and the understanding would be reinforced through labelling places of the number 23.4. The teacher thought the concept of place value was the foundation of learning decimals. In this part, she planned to do a card game in which children would manipulate their digital cards to represent the decimal numbers that the teacher read. Finally, the teaching would be about ordering decimals.

Teacher EN10 planned to build on yesterday's lesson about "percentage of a number and percentage off" to teach the calculation of increasing/decreasing a number by a given percentage. He planned to use football as an example and anticipated it would work catching children's attention.

Chinese reflection on the teaching plan

Teachers CN1, CN2 and CN3, based in the same school, all planned to teach how to order a large amount of data into a table – the first session in Chapter 9 – Statistics and Probability of Mathematics – in the textbook series 4A (pp.70-74). Maths teachers in the same year group of her school always planned lessons together, maintained the same teaching pace and did the same assessment since they were using the same set of textbooks. In addition, they also assigned the same homework every day. Because they shared the same office, it was very common they discussed the results of children's homework while they were marking it.

Teacher CN4 planned to teach Recognising Whole-wan Numbers (认识整万数, see footnotes 5 & 6). The lesson goals were pupils should be able to read and write numbers at the wans level (万级, ibid.) accurately and quickly. **Teacher CN5** prepared the same content as **Teacher CN4**'s as they had usually been doing in the same year group. He also referred to the Teaching Guidance that was published by the provincial textbook publisher and aimed to help teachers plan lessons.

Teacher CN6 usually maintained the same teaching pace as that of **Teachers CN4** and **CN5**, though the lesson observed was about a different topic – constructing bar charts – the second session (pp.75-78) in Chapter 9 – Statistics and Probability. During the lesson planning process, he tried to build the lesson steps on children's prior knowledge to facilitate "positive transfer" as the teacher put it. They had learnt bar charts in which a unit of the Y-axis was equal to five or ten, instead of one. This lesson was about constructing the bar chart where the unit of the Y-axis was one, so the teacher planned to start from the old content and lead pupils gradually to the new.

Teacher CN7 planned to start from the historical story that ancient people represented numbers through tying knots on a string/rope and lead this to the previous knowledge of decimal system. Then, the class would learn to generate the fact that when the number of beads on the thousands' column reached ten, they should put back the ten beads on that column and add a bead to the place higher than the thousands. At that point, the teacher would introduce the wans place (万位, see footnote 5). After that, other three digits at the wans level (万级, ibid.) would appear and be introduced sequentially in the process of the whole class manipulating their abacuses and counting numbers. During the process, there would be reading and writing tasks. In the end, the teacher planned to present real-life pictures which involved numbers at the wans level, such as the picture of the Olympic Stadium in Nanjing.

Teacher CN8 planned to teach the unit on Problem-solving Strategies. Because it was the first time at the primary stage for pupils to learn "strategies" formally, the teacher planned to start from the Crow and the Picher, an Aesop's fable that pupils had previously learnt in their Chinese class. The teacher planned to ask pupils to tell the story and she would then lead them to think about how the crow finally managed to drink the water. It was at this point when the concept "strategy" would be formally introduced. She would then pointed out that, in mathematics, people would also need strategies to solve problems. Because the lesson was to introduce the use of tables to organise information and solve word problems, the teacher would then show a picture of the class's weekly timetable and ask pupils what lesson they would have during a specific slot on a specific day. After that, she would ask why they could so easily find the answer, which would lead to the idea of using tables to get information organised. The main part of the lesson would be following that. The "strategy" would be taught through two exemplary tasks. The key point was to get children realise the importance of distinguishing useful information from useless information when there was more information than needed.

Teacher CN9 planned to review the lesson content prior to the lesson which was about whole-wan and whole-yi numbers (整万数和整亿数[7]). In the revision, she would ask pupils to talk about place value of some given numbers. Then, the main activity would be about numbers that consisted of non-zero numbers in each column at the wans and yis levels (万级和亿级 footnotes 5 and 7) but the lower places did not necessarily contain zeros. The end of the lesson would be a conclusion on what had been taught and learnt.

This subsection shows teachers' reflection upon their teaching plan. Next, we will look at how they feel their lesson plans went.

5.3.2.2 Self-evaluating how the plan went (itvQ5)

English self-evaluation of the implementation of the lesson plan

Teacher EN1 thought his teaching plan was carried out as he expected. **Teacher EN2** was also happy to see the plan went well and said, "It's certainly what was planned to happen."

---

[7] The Yis (亿) Level consists a group of four digits sitting next to the left side (i.e. higher value) of the Wans level (footnote 5). See the English-Chinese corresponding column names below:

English column headings	*Hundred billions*	*Ten billions*	*Billions*	*Hundred millions*
Chinese column headings	Thousand yis 千亿	Hundred yis 百亿	Ten yis 十亿	yis 亿

Whole Yi numbers are numbers that have zero in columns from the units through ten millions. For example, 16 0000 0000 reads as sixteen yi.

Teacher EN3 expressed that he had carried out his plan, but the only thing that surprised him was that pupils were struggling with perimeters of irregular shapes rather than – what he predicted – areas:

> Perhaps the jump to irregular shapes may be a bit too much for some of them, so I took a bit longer on the main [activity]. It didn't keep the exact time of it, but I got all the things that I wanted. The ending was a bit squashed.
>
> (Teacher EN3, 12 March 2013)

Teacher EN4's lesson did not go as she planned, because she did not expect "lots of misconceptions" among pupils. However, she said, "Other than that, all sort of went how I planned it." A misconception she chose to talk about was:

> That first box [the top left cell in the grid method] on the multiple of ten times by the multiple of ten, they were assuming it would always be a thousand – a thousands number. And it isn't always. They were putting in extra zeros onto numbers that shouldn't be. Monday I'll probably look at that again and do a bit more on that, cos they can't make assumptions and also probably do an estimate first, so they know roughly what the answers gonna be.
>
> (Teacher EN4, 9 May 2013)

Teacher EN5 thought that her plan "went OK" and what she planned happened. Nevertheless, she mentioned again about the difficulty in accurately predicting the speed in which this lower maths set might be progressing in a lesson:

> … you always thought, "I wish I got time to get that bit in", because you're never sure how long it's gonna take to recap things, you never know how quickly they gonna remember things they know. What you hope doesn't always happen.
>
> (Teacher EN5, 14 May 2013)

Teacher EN6 planned to do more complicated fractions of a shape but soon found that the majority of the class could not remember what she had taught them four days ago. Then, the planned lesson became a recap:

> They found it trickier than I thought they would. They needed a lot more consolidation than actually, [judging] from the books that they showed me four days ago. They had got it but they got it with me. They needed now to practise it independently. So I had to spend a lot more time on that than I'd first imagined. But, that's just adapting the lesson to meet the children's needs. I found the low ability they moved on to the tricky one – the tricky fraction of a shape.
>
> (Teacher EN6, 1 May 2013)

Teacher EN7 still found some of the majority that she kept on the carpet for the new lesson did not get it, but she was determined to stay optimistic and flexible:

> You (the pupils) don't get it. Let's do some more on the carpet. When I actually sent my top table off, I noticed that they've done exactly the same wrong and that was my fault, cos I didn't say it. So I have to bring those children back to the carpet, say, "This is where we got wrong yesterday. It's my fault, I didn't tell you." I went completely off plan. They got everything, but not in the order. They met everything I wanted them to but it wasn't as quick and effective as it should've been, because I have to go over things again today.
>
> (Teacher EN7, 1 May 2013)

In reality, the content **Teacher EN8** originally aimed to deliver was not fully covered – high-ability children did not get to the point of solving word problems. The teacher thought that the start was good, because children were successfully split into groups of calculating 50%, 10% or 1% of given numbers. Then, the even lower children had spent time working on dividing a number by 10 and dividing it by 100, which the teacher thought going well too. Reflecting upon the levels of tasks she assigned to children for independent work, she thought on the one hand, she might have expected high-ability ones a bit too high and on the other she anticipated low-ability ones would need "a lot of explanation" but they actually moved on to the next level. She said:

> Where I think all went a little bit wrong is those higher children, although I explained it. I didn't think they were necessarily ready to do it on their own, but sometimes they were not good at self-assessing. They are not gonna say, "I don't think I can do it." And then they go back, and they say, "Oh, I can't do it." They come back, which is good. But it's that how long they spent thinking about it before they come and ask for more help.
>
> (Teacher EN8, 1 May 2013)

The quality of assessment is crucial because it points to the next step that the teacher and/or the learner(s) should take. Who have the expertise in evaluating the quality of learning – the teacher or the pupils? To what extent should a teacher rely on children's self-assessment?

Teacher EN9 thought she had stuck to her plan. She seemed particularly surprised to see the low-ability pupil actually progressed very quickly and had a much more solid understanding of what a decimal was. Thus to "adapt things slightly", she tried to get the child working more independently. She also found lots of children understood quite quickly. After they could order sequences of numbers in one decimal place, they quickly figured out how to order numbers to two decimal places. She was delighted to see children progressing so quickly and having a "secure understanding of what a decimal was."

Teacher EN10 thought the lesson went as he planned, since every child in the class made her/his own progress and knew how to increase and decrease a number by a certain percentage. Though the level of the second task was a bit too high, but it was OK, he concluded.

Chinese self-evaluation of the implementation of the lesson plan

About half of Chinese teachers thought their lesson plans went well. **Teachers CN1, CN3, CN4, and CN6** thought they had carried out what they each planned. **Teacher CN2** thought some pupils were very likely to record data inaccurately, which she had not given specific explanations as to why this happened and how to tackle the problem afterwards.

Two teachers thought the plan went well except for tiny flaws. **Teacher CN5** thought the teacher-pupil interaction was frequent but pupil-pupil interaction was not sufficient. Overall, he thought it was OK. **Teacher CN7** thought the lesson went well, except for the fact that she wrote the name of the thousand-wans' place at a later time point than she planned.

Two teachers from School CN-D talked a lot more than did other teachers. **Teacher CN8** pointed out several places where she should have done better. Firstly, she thought she should have shown a very messy list of weekly subjects before showing the class the timetable, thus to form a bigger contrast between messy information and well-organised information so that children could see the beauty of organising information into neat tables. Secondly, she regretted that, in the second example, it would be better if she let the pupils figure it out themselves rather than she suggested pupils to use arrows instead of tables to illustrate relationships between known and unknown numbers. She did so because time was not enough. Thirdly, she thought more emphasis should have been put on to pupils' understanding of quantities and the relationships between them. Nonetheless, she thought that it was both good and bad to make a lesson perfect, because she was used to reflecting upon her teaching and planned to make it up in the next lesson. **Teacher CN9** thought the lesson was carried out as she expected, but there were quite a few points that she thought could be better. For example, the beginning of the lesson was about the distance between the Sun and the Earth, and the type of the number was what the lesson intended to address. She thought it was an inappropriate example which would be better to appear later, but luckily pupils still managed to read it correctly.

Summary

Almost all teachers thought they had carried out the lessons as planned. At the same time, there were also some teachers from both countries regretting for things that they had not anticipated or planned well, but the types of things on the regretting list were quite different. English teachers were likely to find out the difficulty level of content and/or the pace of teaching they foresaw did not match the reality, which explains why they had to give up plans in a varying degree to adapt to the reality. On the contrary, Chinese teachers who regretted were all happy with the kinds of things that English teachers mourned for, and what they actually regretted were the imperfectness of certain small steps.

In the following subsection, we will hear teachers' comments on their own lessons and the effectiveness of their teaching whilst and after viewing the lesson video.

5.3.2.3 Self-comments on the lesson & its effectiveness (itvQ6)

English self-comments on the observed lesson

Teacher EN1 thought it was important to show the class the lesson targets, steps of Success Criteria, at the beginning of the lesson, because children would know "what they are going to do today". He thought it was also important to ask pupils to talk about why they were going to learn times tables. Then, he emphasised the importance of using the quiz as a screening/differentiating tool in which those who felt difficult could leave the carpet at any point and go to the TA's table for support. In his view, the process of the lesson gave children opportunities of "self-assessing and self-beating" and allowed them to self-assess where they were at the end of the lesson and to plan where they were going to work on the next day through the "traffic light" activity. He saw this as an effective lesson in terms of what he wanted children to learn and "in terms of retaining their times tables."

Watching the Q & A at the beginning of the lesson, **Teacher EN2** was pleased to see that children were "very keen to share ideas" on why they should learn times tables. At the beginning, pupils were talking about why they needed to practice times tables, the teacher called a boy's name and paused for a while. In the interview, the teacher explained that she picked the boy up because he was the one "who won't concentrate". While other children were about to practise and self-assess times tables, several pupils were asked to join the teacher on the carpet. The teacher said that they needed extra support with their number bonds. Overall, she thought that the pupils were all on task, they were all achieving and making progress, they knew where they could get things and how to move on, and therefore the lesson was effective.

Teacher EN3 found that his pupils were all on task and that they were quite quickly getting on with the self-assessment. However, he found the middle group might need more support and he would work with them tomorrow on the perimeters with missing edges. He regretted again that the ending was a bit rush and decided to end more formally the next day. To summarise, he thought the interaction, paired discussion, paces and "all the elements" were there, so the lesson was effective, but maybe he had pitched it a little bit higher.

Teacher EN4 thought her middle set was a hard maths group where there were "a lot of behaviourial children". Looking at the boy A raising his hand eager to talk, she said A is quite difficult, B is quite difficult, C, D and E – they were quite difficult. The main activity went "longer than usual". Overall, she thought the lesson was effective because the pupils had recapped the grid method:

> They aren't perfect at using the method yet, but I think they know where's gone wrong, so I know what it is I need to do next to move their learning on.
>
> (Teacher EN4, 9 May 2013)

Teacher EN5 introduced a bit more about the lower set while watching the lesson video. There were four SEN children, including a Down syndrome girl and an Autistic boy. The Autistic boy was actually good at maths. To meet SEN, two TAs were working alongside her and sometimes she also got a third TA. The teacher pointed at a girl on the screen, saying that the pupil was really struggling in rounding down decimals – "she got too far down". She talked about her strategy – giving them the first chance to do it to see how they got on. When the video was showing a girl covering her book with the hands to prevent her partner seeing her work, the teacher laughed out loud. She thought pupils were good at staying on task. To summarise, she thought this was more of a recap lesson and did not clearly comment on the effectiveness of it.

Teacher EN6 explained the purpose of her action at some points whilst viewing the lesson video. Firstly, on the carpet, she started to recap what had been taught four days ago. During the process, she was assessing children and sending low-ability ones away to independent work. The Success Criteria were broken down to steps and children were divided into three sets. By looking at the children's work on their WBs, the teacher made decisions as to who could do it and were therefore ready to leave the carpet. When everyone was working independently, she explained that her purpose of walking around was to check whether anyone was struggling and to bring her/him back for extra inputs if she/he was. She emphasised the importance of asking more how-questions, continuous assessment and making sure she knew where the pupils were. She said she would send both high- and low-ability children to the TA for more advanced learning or extra support. From time to time, she would try and talk to the TA to see whether everything went well, because the TA was not a qualified teacher after all. She highlighted the "traffic light" activity at the end of the lesson where children coloured the steps they had reached (green = I got it, yellow = a bit more help, red = I didn't get it). To conclude, she thought the lesson was effective.

Teacher EN7 thought all pupils were on task and well motivated. She said she would always keep an extra slide for those who finished earlier, and that the Success Criteria also helped tell pupils what was next if the teacher was busy. The teacher mentioned the use of a game to summarise and consolidate the knowledge at the end of a lesson. Overall, she thought the lesson was effective because she had spotted what misconceptions the children had.

Teacher EN8 thought overall the lesson was effective because the children were appropriately differentiated and they all had the enjoyment of mathematics.

Teacher EN9 pointed out it was good to spend longer time – longer than usual – on the carpet for the input. She thought if she had not done the long input, the children would not have the "secure understanding" which helped them progress and speed through the learning. Overall, she thought the lesson was effective because every child, by the end of the lesson, could tell what a decimal was in terms of place value.

Teaching the top set of Year 5, **Teacher EN10** appreciated their level of maths and thought some of his pupils were much faster than him. At the beginning, he regretted and said that he should have asked the pupils to talk to the person next to them. He pointed at two children and said they were quite good at maths – like mathematicians. At point 10:20 on the video timeline, he said that he should have done some calculation instead of talking. He further mentioned that he usually prepared extra exercises for high-ability children. Overall, the lesson was regarded as effective because all children had made progress.

Chinese self-comments on the observed lesson

Teacher CN1 paused at a point and said that part of the activity was aimed to enhance children's ability in interpreting data. When pupils were asked to discuss with their "desk-mates", the teacher said her purpose was to allow them – particularly those who were a bit shy or unconfident – to take the opportunity to talk and listen to others' views before they were called to share ideas with the whole class. The best way of ordering data was collectively found through teacher-guided whole-class discussions and then the teacher asked pupils to try the agreed strategy with a task on their textbooks. They did it very quickly. Immediately after that, the teacher suggested a collaborative activity between three parties – the teacher, a volunteer from the class and the rest of the class. The collaboration was to redo the task through three actions: (1) the volunteer pointing at each entry of the raw data on the screen (PowerPoint slide) with a long wooden ruler, (2) the rest of the class reading loudly the code of the category that the specific data entry fell in, and (3) the teacher jotting down one of the five strokes of the character 正 (Chinese tally marks) in the corresponding cell (height range) of the table. When the lesson video went to that point, the teacher talked about her intention of including this "three-party" collaboration:

> This process involves me writing on the board, a boy pointing at the data and the class telling me where I should jot down a stroke. I did this because I was still not very sure if everyone had grasped the method solidly. I had to consolidate at this point, because teaching examples is the most important part of a lesson. I must make sure every pupil clearly understand the method through the teaching and learning of *one* example. I can't patch up here and there later when misunderstandings are already there. This process gave every pupil a second chance to judge which entry went into which category. After this, they should have a clearer understanding of this method.
>
> (Teacher CN1, 11 December 2012)

Teacher CN1 also mentioned several places where she thought she should have done better. For example, along with the process of demonstrating a table, she thought she should have presented corresponding notes on the board at a right time accordingly. “Because we should leave on the board a complete set of 板书 (Chinese pronounced as *banshu,* meaning notes on the board in English) at the end of each lesson”, said the teacher.

Like **Teacher CN1**, all other Chinese teachers thought their lessons went effectively, but they also found small things that they regretted for.

Teacher CN2 thought she should have emphasised that in the final exercise there was not a table for ordering data as in the exemplary task. If she did, pupils would all know the only table provided must be filled in with numbers rather than Chinese tally marks “正”. She also said that the lesson would have been better if she left more time for connecting data to environmental protections.

Teacher CN3 thought the middle part of the lesson could have been quicker. She talked about her ideas of showing pupils’ work via the projector to the class as a conclusion to every task and thought it worked very well.

Teacher CN4 thought the reading and writing methods of numbers at the Wans level was well structured, but the Level-separating Line[8] was not thoroughly taught.

[8] A Level-separating Line functions similarly as a comma (e.g. in 10,000) in the Western numerical system. It is used informally as an aid for children to read large numbers easily.

Teacher CN5 felt all his targets had been achieved but there was still a bit of dissatisfaction in terms of meeting high-ability children's needs.

Teacher CN6 thought the lesson went well but not as perfectly as he imagined.

Teacher CN7 regretted for forgetting to introduce the Western numeral system which would separate numbers by three digits and ask pupils to compare that with the Chinese numeral system. She also found the ending was a bit rush, but she was satisfied with the fact that the lesson put sufficient emphasis on place value and the Level-separating Line.

Teacher CN8 thought it was good to use the story of the Crow and the Pitcher at the beginning of the lesson which gave the children a vivid sense of what the word "strategy" meant. **Teacher CN9** thought she should have stressed more on reading and writing numbers at the yis (亿) level (footnote 7), but it was good to stress the concept of the Level-separating Line.

Summary

Almost all teachers agreed that their lessons were effective overall. Two English teachers, **EN1** and **EN2**, both from School EN-A did not find any particular elements of the lessons that they were not satisfied with, whereas the rest of the English teachers and all Chinese teachers each expressed a varying number of things that they thought were not arranged well enough. The majority of teachers showed critical thinking upon their own practices, which is an essential step towards the improvement of teaching. Besides, Chinese teachers were more concerned with their ways of dealing with specific points of knowledge, while English teachers cared more about their approaches to grouping children and differentiating teaching and learning.

Next section is about the last of the three interview components which consists of two questions – international awareness and flexibility to change.

5.3.3 International awareness and flexibility to change

In this part of the interview, teachers were each firstly asked to describe their imagination or impression of the mathematics teaching to the same age group in the other country – in England for Chinese teachers or in China for English teachers. Secondly, they were asked to talk about their flexibility to change in their future practices.

5.3.3.1 Imagine a maths lesson to the same age in the other country (itvQ7)

English imagination of a Chinese lesson

With regard to a maths lesson to the same age group in China, **Teacher EN1** imagined more advanced mathematical content, pupils facing forwards, more individual learning and self-support and more skills based rather than the application.

Teacher EN2 thought Chinese lessons would be "much more formalised" and "much more structured". Pupils would be given things to work on and they would be "working with their times tables". The lessons would be "more adult-led than child-led", and their desks should be set in lines, with children facing the front.

Teacher EN3 imagined a Chinese lesson would be more structured, with more abled and independent learners and less paired discussion. The children would be sitting in rows, and the teacher would be at the front "more like a university lecturer".

Teacher EN4 thought in China the maths lesson would be very structured, and all children would be sitting in rows, working. She imagined "their teacher at the front teaching a skill and everyone getting it straightaway."

Teacher EN5 thought Chinese lessons would be more about rote learning and a lot stricter, with everyone on task and sitting in rows.

Teacher EN6 said her imagination might be a bit stereotypes – a class with desks facing the front, children not in groups and without free movement, whole class teaching, teacher at the front, and using textbooks.

Teacher EN7 imagined that the Chinese lesson would have one or two children per desk, the class would use textbooks, and the teacher would give commands, such as "Do this! Go!" She took her impression on a French lesson that she had seen before as a similar example and said that there was very much teacher talk at the front.

Teacher EN8 thought children in the Chinese class would be sitting at individual desks working independently, the teacher standing at the front and explaining what they need to do. She thought less of children would be able to come to the teacher to talk on the carpet and all of them would be:

> very driven, very motivated to work really hard. There's more of motivation rather than having to make the lesson really fun and things going on. I think they'll be quite motivated.
>
> (Teacher EN8, 1 May 2013)

Teacher EN9 also imagined the Chinese class would have desks in rows and textbooks. There would not be much of differentiation. She did not imagine children would be much challenged, but they had to adapt to the teaching pace. If they fell below or above, the lesson just went on as normal, she imagined.

Teacher EN10 had the experience of working in South Korea as a secondary English teacher for four years and he imagined that Chinese classes would be similar – children sitting individually and teacher lecturing. He also mentioned that Japan had been doing things in a similar way.

Summary

Imagining features of mathematics lessons in China, English teachers had significant agreement on two aspects: (1) pupils sitting in rows, sitting individually or in pairs, facing the front, and/or not allowed to move freely (8 out of 10 teachers) and (2) teacher-led and/or teacher at the front (7 out of 10). Three teachers reckoned Chinese would be using textbooks in classrooms; of six teachers, every two shared one of the following views: (1) more structured lesson, (2) more advanced/challenging mathematical content and (3) children well motivated, on task and hard-working.

Chinese imagination of an English lesson

Teacher CN1 imagined the English teaching style based on her impression on American education. She had a friend in Shanghai who recently relocated in the US and had her six-year-old attending a local school. Through her friend's experience, she knew in an American class for this age there would be no more than twenty children with several teachers who would give them elaborate attention and care, holding loose standards. She thought English classes might be similar to that and she further counter compared that with Chinese classes:

> We are holding higher standards in discipline in order to meet our goals in terms of the effectiveness and pace of teaching and learning. I think if the Western classroom can be that loose and relaxing, their curriculum

> must be much easier than ours, their pace would be much slower and the expectations from parents might also be much lower.
> (Teacher CN1, 11 December 2012)

Teacher CN2 thought a maths class in England would have pupils sitting at round tables, seek to make them happy and free and give much attention to their behaviours and habits.

Teacher CN3 thought England's maths lessons would be more open, children would be more independent and individuals would be more respected.

Teacher CN4 imagined more freedom in the English classroom and more relaxed sitting postures of children. The class size would be around 20, and children would sit in circled groups and be free to choose which group they would like to join. If children did not get what the teacher taught, they could ask the teacher teach them again.

Teacher CN5 imagined that English children would be more relaxed. The most important thing was to create a good atmosphere rather than get children to know everything. It was already okay if the children had a little bit of understanding. English schools did not care a lot about what problems the pupils could solve in the end.

Teacher CN6 thought English lessons would be freer and looser, unlike the Chinese lessons which were more formal and stricter.

Teacher CN7 thought the English lesson would be simpler and easier than the Chinese lesson. She seemed to be more familiar with the American style and thought the American lesson would be more open and passionate. The English lesson might be slightly more formal than the American lesson, but it would be more flexible and more related to children's life experience than the Chinese lesson. The desks would be set into groups and the teacher might not be at the front.

Teacher CN8 also thought the English lesson would offer children more freedom but at the same time lower standards.

Teacher CN9 imagined the English lesson would be more open than the Chinese one and teachers could be more flexible. Children could sit in whatever postures they would like to. They would be grouped into circles and could talk about their ideas at any points in time.

Summary

Two thirds (n = 6) of Chinese teachers shared the view that English maths lessons would be more relaxing, set looser standards and offer children more freedom. The next most popular view (from 4 out of 9 teachers) was children would be sitting around tables in groups. Three Chinese teachers thought the English lesson content would be much easier, and three others imagined it would be quite open.

Two thought English children would get more individual care, and two thought the class size would be smaller – around 20.

It seems that teachers from two countries imagined more about the role of the teacher, the way children might be grouped in class, and the level of mathematics they were possibly expected to learn. In the next section, the focus will be on teachers' flexibility to change in their future practices.

5.3.3.2 Flexibility to change in future (itvQ8)

English teachers' flexibility to change

All English teachers expressed that they were open to change once potential teaching approaches were identified.

Teacher EN1 said that he was "always trying new things". **Teacher EN2**, in her third year of teaching, said that she probably had changed a lot from her "first year to now" and had recently been suggested to borrow the approach of starting with a ten-minute mental maths task from another school after visiting that school. She said:

> It's always change. I think you have to be, as a teacher…. You pick up on things that are good for the children.
>
> (Teacher EN1, 26 February 2013)

Teacher EN3 cherished each new idea he encountered and said that his teaching changed a lot during the last two to three years. **Teacher EN5** also said that she was always open to change. Likewise, **Teacher EN4** from the same school said:

> Yeah. [I'm changing] all the time. I mean we've changed so much recently. If you'd come to the school a year ago, you would've seen completely different maths lessons. The cutaway technique is quite new. The Success Criteria is new. The self-assessing is new. We're all open to change and to what's best for the children, cos that's what gets you to the effective lesson, isn't it?
>
> (Teacher EN4, 12 March 2013)

Teacher EN6 said she was always looking for ways to improve her teaching and, as the Year-5 leader of the School EN-C, she expressed that being open to change was the culture of the school:

> [At the school,] it's OK to get things wrong and it's OK to find a better way of doing something.
>
> (Teacher EN6, 1 May 2013)

Teacher EN7 said that she would definitely change if it could meet the needs of children and she further talked about her development by putting herself in the context of the school:

> We're very, at this school particular… We've got a very set style with 'Cutaway' and you know, having the whole class and then other groups still showing off now and again. To be honest, cos I started my teaching here, it's all I know, but I know that there're other ways of doing it. So I think, when I do move to another school, it'll be hard [to adapt without changing]. … but I know that's not the only way that we can teach, definitely. But, I find it hard not to do it. I really would [change].
>
> (Teacher EN7, 1 May 2013)

Teacher EN8, as a first year teacher, said that she would definitely like to change. Looking back to the past several months, she thought her teaching had already changed a lot. She said, "Whenever I got some new ideas, I would try it out and see how it works."

Teacher EN9 appreciated the opportunity to change and thought she already adapted since she was not always teaching in the same way. About her change influenced by the school culture, she expressed:

> I see the benefits of doing it. I find it easier to teach. I find that the children get more out of it, although it's more time-consuming. There are more resources [and] more to think about. The outcome is so much better. If there's another option to try something else, so long as the outcome is good, the children is learning, then, yeah, I would give it a go.
> (Teacher EN9, 8 May 2013)

Teacher EN10 had just come to the school recently after teaching English in Korea. He thought he had already adapted to the teaching style "here" very quickly, though the style of the school was new to him.

Chinese teachers' flexibility to change

Chinese teachers each expressed a different degree of openness to change, but the majority (n=6) of them also thought that there would be something hard or impossible to change and therefore a complete transplant of new approaches was impossible.

Teacher CN1 showed her openness to change, but she also pointed out that there were some aspects that were hardly changeable. She said one had to firstly change the whole society, the curriculum, the required teaching pace, parental expectations and school management and finally some change could be hopefully brought into the classrooms.

Teacher CN2 said she would be open to change. As a first year teacher, she was always looking for new approaches.

Teacher CN3 was willing to learn new methods that might work in her class. When talking about the possibility of taking up English approaches of teaching, she said it was very unlikely for the Chinese schools to learn from the West. She said:

> Some say America is children's heaven, but many Chinese parents find their kids learn too little in American schools. They take the Chinese textbooks with them so they can teach their kids at home after they immigrate to the US. They don't feel very confident about the quality of American education.
>
> (Teacher CN3, 11 December 2012)

She further pointed out that the basic issue was that both countries had different educational systems and this was why it might not work if the Chinese borrowed Western methods. She was also concerned with the larger class size in Chinese schools and said the Western approaches would not work with over 40 children in a class.

Teacher CN4 would like to change slightly but thought it impossible to transform his style completely into others'.

Teacher CN5 expressed his openness to change and said he had been consistently learning from others.

Teacher CN6 said that his teaching beliefs might change a little after seeing new approaches of teaching, but his teaching practice would not change immediately.

Teacher CN7 would like to learn useful methods or experiences from other teachers either home or abroad. Nevertheless, she thought it would not work to completely copy others' approaches and abandon every bit of her own methods. She thought both cultures could learn from each other's good aspects whilst respecting each other's differences.

Teacher CN8 said she was open to change and had been changing her style constantly, but she thought it was important to keep a critical eye before making any change.

Teacher CN9 said she was very possible to change. As a new teacher, she had not stabilised her own beliefs, so she was always open to new ideas and approaches.

Summary

Nineteen teachers all said they were open to change, but there were apparent national trends. While teachers from England were more flexible and used to change, two thirds of teachers from China were more concerned about issues that might prevent significant and/or fast change, such as the differences of cultures and educational systems. Of the three Chinese teachers who had not mentioned any obstacles to change, two were first-year teachers showing their willingness to actively learn anything new that could enrich their practices immediately. To this point, all interview data have been presented and discussed, which sheds light on the quality of teaching from the angle of individual teachers' beliefs and self-perceptions. The next main section will further reveal teachers' collective views on two lessons, one from England and the other from China, in the focus groups regarding the effectiveness of mathematics teaching.

5.4 Focus groups with teachers (M6)

In each focus group, teachers were shown two lessons, EN2 and CN7. For each of the two lessons, firstly the subject content and the lesson process will be briefly described; secondly, native and foreign colleagues' views will be presented and interpreted. Before dipping into detail, it should be pointed out that the length of each subsection was solely determined by the amount of information that each strand of data had to offer. For example, the focused English lesson did not consist of lots of teaching events, nor much change of classroom activities, so that part was relatively shorter than the part on the focused Chinese lesson which had considerably more events and thus demanded more words to describe.

5.4.1 The focused English lesson: Content & process

The subject content of the focused English lesson

Given the fact that children had learnt times tables in the previous year, Lesson EN2 was aimed to consolidate children's times tables with an extra attention to a few low-ability pupils' 10 and 20 number bonds.

The process of the focused English lesson

EN2 Component-1: Reasons & targets to learn

This lesson was the second in a sequence of lessons on the same content – Times Tables. At the beginning, Teacher EN2 spent 7 minutes on whole-class interaction in which the teacher firstly asked children to talk about the purpose of learning times tables and then showed them the Steps of Success Criteria on the IWB (Figure 5.10).

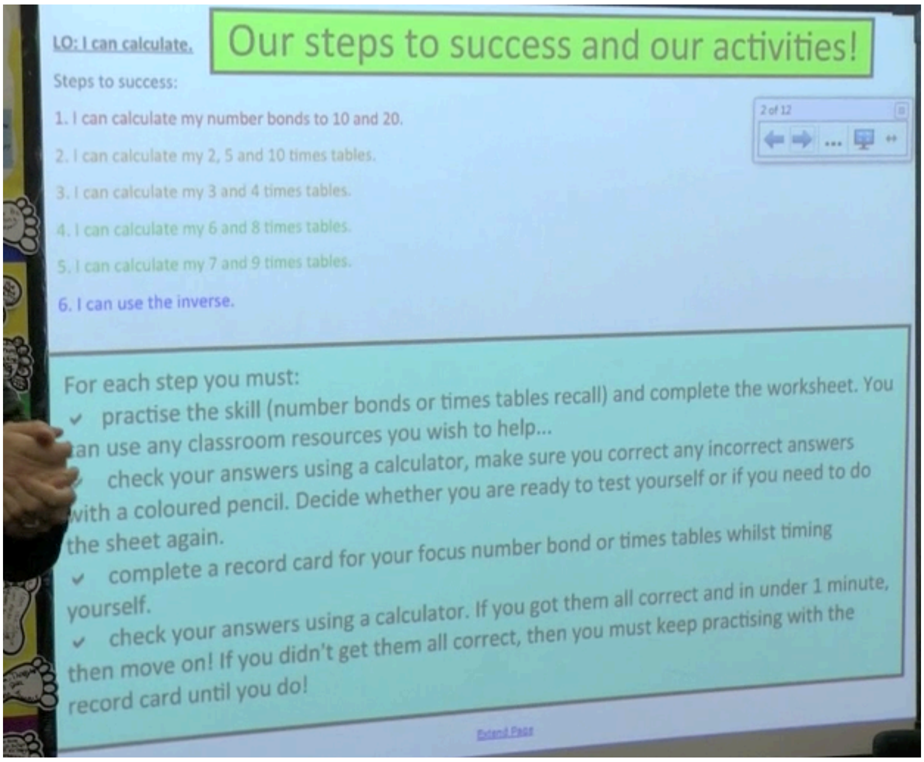


Figure 5.10 Lesson objectives of Lesson EN2

Figure 5.11 Differentiated worksheets in Lesson EN2

EN2 Component-2: Independent work & self-assessment

Then pupils were sent to independent work on coloured practice sheets (see Figure 5.11). Such independent work lasted 40 minutes in which children selftimed their calculations, self-assessed their answers with calculators and progressed to the next-higher-level worksheet if they found they took no more than a minute to complete the prior-level sheet correctly. While most pupils were working independently, the teacher and the TA were each working with a very small number of low-ability pupils.

EN2 Component-3: Self-assessing goals achieved and to be achieved

During the last four minutes of the lesson, the teacher asked pupils to read through the learning targets (i.e. Steps of Success Criteria), reflect upon their work, mark what they had achieved that day and plan what they would work on the next day. The lesson ended here, and the teacher started an English lesson without a break.

Next section presents various opinions from teachers in the focus groups organised within each country regarding the quality and effectiveness of the focused English lesson.

5.4.2 The focused English lesson: Colleagues' views

In this section, first come English teachers' voices, and then Chinese teachers'. As introduced in previous chapters, there are two focus groups in each country, which makes a total of four focus groups: EN-FG1 (n=5), EN-FG2 (n=5), CN-FG1 (n=6) and CN-FG2 (n=6-3)[9].

The English lesson – native colleagues' views

Five English teachers in focus group 1 (EN-FG1) thought that Teacher EN2's lesson was not effective:

> They (the pupils) were just practising skills to make them quicker…She (i.e. Teacher EN2) was there facilitating them doing sheets – that's how I saw it … There was a lot, at the beginning, of teacher talking to the whole class. But not an input teach[ing]. It wasn't giving them anything.
>
> (EN-FG1, 22 May 2013)

In the other English group which included Teacher EN2, teachers agreed during the group discussion time that the lesson offered more differentiation to children and were more pupil-led than the Chinese lesson. Nonetheless, they later on expressed their actual opinions about the English lesson to the researcher personally in their school:

> The teacher only spent ten minutes out of about one hour interacting with the whole class and there was no content going on. ... It was unbelievable that the English teacher took a whole lesson on times tables. The content was unbelievably easy. That English school should have set by maths….
>
> (EN-FG2, 06 June 2013)

[9] Initially six teachers took part in CN-FG2, but three opted out later (as explained in Sections 3.3.3 and 3.4.6).

Overall, all native colleagues but one (Teacher EN1) genuinely considered that the lesson EN2 was ineffective, time-consuming and short of teacher input. It would be interesting to hear what foreign colleagues reckon about the same lesson.

The English lesson – **foreign** colleagues' views

Most Chinese teachers thought the English lesson offered children unlimited opportunities to explore their learning, to choose the level of worksheets that they felt they could do and to self-assess their learning, which were quite different from the Chinese maths class. They pointed out:

> The (English) lesson was more relaxing and individualised, and the children more independent. … The teacher did not need much time to prepare lessons (i.e. do daily plans)….
>
> (CN-FG1, 25 December 2012)

On the other hand, they also thought the English lesson had low expectation of children's calculation ability and had less content delivered over 50 minutes. Some teachers argued:

> If the final exam paper were the same to every pupil, then differentiated learning goals would mean nothing and particularly unfair to those low-ability children, because they would get worse and worse, being taught like that.
>
> (CN-FG1, 25 December 2012)

Teacher CN4 thought that those children in the class were probably of the level of Grade 2 in China (i.e. Year 3 in England, meaning two years behind China). Discovering that the teacher spent quite a long time giving guidance to one pupil, some teachers talked to each other that none of Chinese teachers would have done that because they had to get the content grasped by the majority rather than a single child over a limited amount of lesson time. Teacher CN8 said:

> This lesson was very time-consuming and thus not effective and it was impossible for Chinese teachers to adopt this approach while maintaining the teaching pace.
>
> (CN-FG2, 26 December 2012)

Nonetheless, the majority of teachers agreed that each country had its own standards, so depending on the English standards this lesson might be effective.

The following two sections show the content and process of the focused Chinese lesson (Section 5.4.3) and the collective views of each country's teachers on the focused Chinese lesson (Section 5.4.4).

5.4.3 The focused Chinese lesson: Content & process

The subject content of the focused Chinese lesson

The topic of the lesson was about numbers from wans (Chinese equivalent place name for ten thousands) to thousand wans (Chinese equivalent place name for ten millions, see footnotes 5 and 6) and their place value. The aim of this lesson was to introduce the names of these four places and how to accurately read and write numbers that fell into this range. As aforementioned, the Chinese numeral system is slightly different from that of the West in that the former separates numbers by four digits whereas the latter by three. Thus, one place up the thousands place is given a Chinese name called wan (万) which is the equivalent of ten thousands in English. For example, the number 10000 is read as “one wan”.

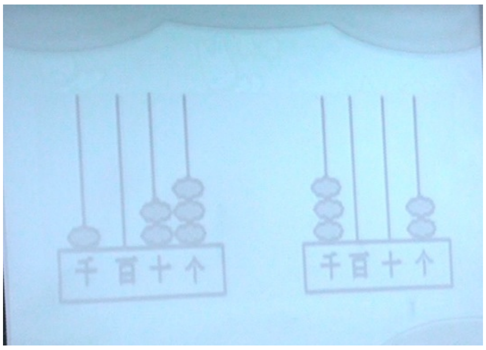
The process of the focused Chinese lesson

CN7 Component-1: Leading in

The teacher started with the historical story of the way ancient people counted tens, hundreds and thousands by tying knots of different sizes. This story led to the idea of base ten. She then pointed out today’s lesson was about “recognising numbers (认数)” while writing down this title on the board.

CN7 Component-2: Reviewing four places (units to thousands)

At the video time point 00:02:16, the teacher asked pupils the name of each number place, and as they calling out the column headings she wrote units (个), tens (十), hundreds (百) and thousands (千) on the board. Then the teacher showed a slide on which two numbers were represented on abacuses; through Q & A, the class reflected upon the reading and writing methods of the two numbers (Figure 5.12).



Translation: 千 = thousands, 百 = hundreds, 十 = tens, 个 = units.

Figure 5.12 Numbers represented on abacuses in Lesson CN7

CN7 Component-3: Introduce new digits

At 00:04:14, the teacher suggested the class to manipulate their abacuses together with her to represent numbers thousand by thousand starting from 5000. Every time after adding a bead onto the thousands, there would be a round of Q & A about how to read and write the number and what the beads meant, for example, children would say that six beads on the thousands meant six thousands. After 9000, 1 0000 emerged, at which point the new place – wans (万 as aforementioned – Chinese equivalent of ten thousands) – was introduced, and the teacher wrote the column heading to the left of the thousands on the board. The activity continued until the number reached a hundred wan (100 0000).

CN7 Component-4: Generate writing and reading methods

At 00:17:36, the teacher showed six numbers on a new slide (Figure 5.13) and initiated a round of Q & A about the thinking method for reading the numbers correctly. The most agreed answer was to count four zeros from the units up through higher places, prepare to read the group of four zeros as one character *wan* (ten thousand) and read loud the remaining number on the left hand side firstly and the character wan secondly.

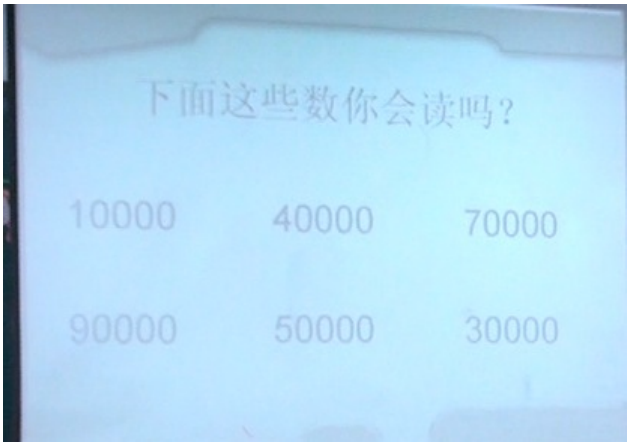


Figure 5.13 Task on a slide from Lesson CN7 (Component-4)

At 00:18:25, the teacher turned to the next slide which showed a picture and a description which read "Nanjing Olympic Stadium had a total of 6 0000 seats." She asked a pupil to read it for the class. Then, another slide came, asking pupils to write two numbers according to their readings in Chinese. Once done, the teacher asked pupils to check the answers with their desk-mates and talk to each other the writing methods they used. To conclude this task, the teacher initiated anther round of Q & A for individuals to share with the whole class their methods.

CN7 Component-5: Generate the idea of the Level-separating Line

At 00:22:34, the teacher showed a slide on which four numbers were given. Then she posed two questions:

1) Could you read these numbers?

2) Are there any methods that could help us read the numbers after a quick glance?

Immediately, pupils started to volunteer to answer the questions. At the moment when a pupil pointed out the idea of grouping numbers by four digits, the teacher spotted an opportunity to introduce the Chinese numeral system where every group of four digits form a level. She then wrote the names of two levels: the units level (个级 from units to thousands) and the wans level (万级 from wans (ten thousands) to thousand wans (ten millions)) (Figure 5.14). The idea of the Level-separating Line (分级线，footnote 8) was introduced following a relevant answer from a pupil. The teacher then drew the red dotted line on the board between the units level and wans level.

Immediately after this, she asked the class, “How shall we divide this number 50000 with the Level-separating Line?” Children answered the questions randomly. When some pupils said it should start from the units, the teachers asked the whole class, “Counting from the units, how many places up?” The class answered, “Four places!” The teachers asked, “Then what?” The class answered, “Draw a Level-separating Line.”

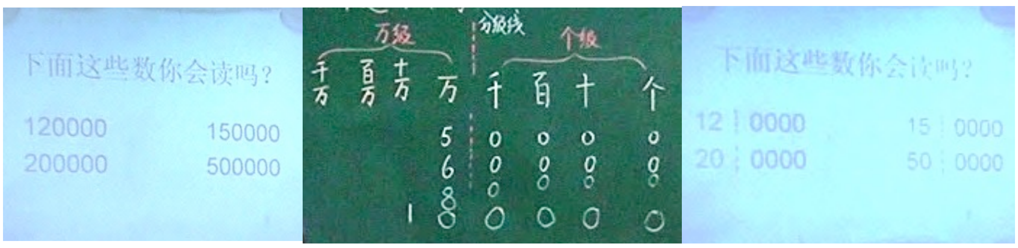


Figure 5.14 Screenshots from Lesson CN7 (Component-5)

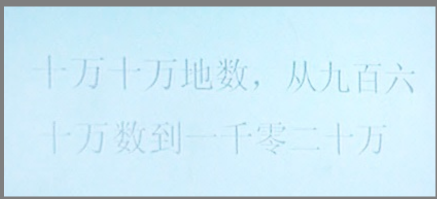
On the board, the teacher drew a dotted line between 5 and the 0 next to 5 in 50000 and asked, "Once we draw the Level-separating Line, is it easy to read the number?" Pupils responded loudly, "Easy!" "How to read it?" asked the teacher. "Five wan!" While children were reading "five", the teacher simultaneously pointed at the number 5; while they were reading "wan", she quickly pointed across the remaining four zeros from the thousands down to the units. She tried to simultaneously complement her body language with children's reading. After this, the method of drawing the line and then reading 6 0000 was carried out similarly but faster. At this point, the teacher suggested pupils to choose three numbers on their textbooks, try out the method independently and then discuss their methods with desk-mates.

CN7 Component-6: Represent numbers with the Level-separating Line in mind

At 00:29:56, the teacher showed another slide (Figure 5.15) asking the class to represent and count a sequence of continuous numbers on their abacuses, and then write the numbers on their workbooks. At 00:33:53, the teacher asked whether the Level-separating Line did help in writing numbers and, if yes, how. Individuals were asked to talk about their methods.

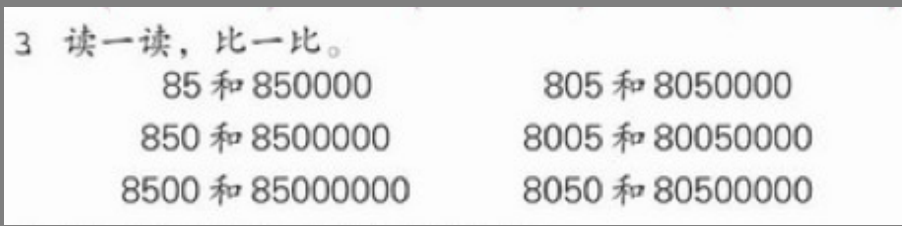
CN7 Component-7: Reviewing and consolidating the lesson content

At 00:30:04, the teacher pointed at the title of the lesson again and asked the class to look carefully for the similarities between those numbers shown on the slide (Figure 5.16).



Ten wan (a hundred thousand) by ten wan, count loudly from 9600000 to 10200000.

Figure 5.15 Task on a slide from Lesson CN7 (Component-6)



Translation (Chinese → English)

读一读，比一比。→ *Read & compare numbers.* 和 → *and*

Figure 5.16 Task on a slide from Lesson CN7 (Component-7)

After a round of Q & A, at 00:39:22, pupils were suggested to turn to page 87 in the textbook and work on problem #3. After a few seconds, the teacher said she found some pupils were all doing the same thing. She then called a girl to tell the class what it was, and the girl said she was drawing the Level-separating Line. The teacher gave the girl a swift praise and recommended the method to the class.

Noticing that the class were all done, the teacher initiated a round of Q & A for individuals to stand up to read the numbers comparatively: e.g. eighty-five (85) vs eighty-five wan (85 0000), eight hundred and fifty (850) vs eight hundred and fifty wan (850 0000), and so on. Another round of Q & A concluded the reading and writing methods for whole wan numbers[10] at the wans level. At 00:43:03, the teacher initiated the last round of Q & A regarding the question "what have you learnt today?" The lesson ended one and a half minutes later.

Next subsection will present teachers' views on this focused Chinese lesson collected through the four focus groups across the two countries.

5.4.4 The focused Chinese lesson: Colleagues' views

The Chinese lesson – native colleagues' views

Chinese teachers across two groups came up with the agreements that this lesson had both a clear structure and logical and coherent connections between sections of the lesson, that the correct rate of children's work would be very high for this lesson and that the lesson represented what the local teachers had been doing in the class. Teachers tended to dig deep into specific steps of the lesson. For instance, Teacher CN2 thought the lesson was taught with great detail and clarity, but the lesson content was not difficult so the teacher should not have allocated long time for abacus manipulations and drawing Level-separating Lines. Teacher

[10] This refers to numbers whose last four digits (from units up to thousands) are exactly zeros, e.g. 4 0000.

CN6 however noted that the reason the teacher spent longer time on the learning of the Level-separating Line might be that the teacher was trying to give the children more time to explore and find the usefulness of drawing the Level-separating Line and consolidate their knowledge of place value. He thought it seemed helpful in developing children's ability.

Weaknesses of the lesson were also discussed at times. For example, in CN-FG1, while acknowledging the lesson was effective in achieving the knowledge targets, Teacher CN1 also argued that the lesson did not offer enough opportunities for children to explore the knowledge. In CN-FG2, Teacher CN8 suggested that the teacher might consider replace the physical abacuses with practice sheets on which children could draw circles to represent beads. With this method, children would be able to gradually develop their abstract-thinking ability by relying less upon representing numbers with physical objects, given the fact that they were progressing towards senior years. Teacher CN7 tended to turn a critical eye on her own lesson. In CN-FG2, she volunteered to be the first to talk and pointed out several of her mistakes in the lesson: (1) mistakenly writing the word 计 (counting) as its homophone 记 (recording/writing); (2) writing the place of *thousand wans* (千万 ten millions) on the board at a time which otherwise could have been a little earlier. She later on added that she should have introduced to the children the differences between the Western (3 digits) and Chinese (4 digits) ways of separating digits.

The Chinese lesson – foreign colleagues' views

Regarding the negative aspects of the Chinese lesson, FG-EN1 teachers offered brief comments regarding the lack of differentiation and the teacher standing at the front and they then went on discussing about its positive points for a long time. They were impressed by the clarity of the lesson, children's motivation, children's respect of teachers and their behaviours.

> The clarity of her explanation in the context was phenomenal. It was so clear. She had the visual of the abacuses that every individual child was doing. She had the visual on the board of the numbers moving. She had children coming up. She had the context. It's very simple, but it's really effective.
> …
> Such motivation! Such drive! In the English lesson, it was the teacher enthusing the children by setting tasks and motivating them by using the timer to beat their time. There is more responsibility on the part of children [in the Chinese lesson]. In England, it's all on the teacher's.
> …
> These children they come to the classroom, [and] they are so motivated to learn. Our children, if we stand in front of the classroom, it would be so difficult to motivate them. You know, we came in one day and we dressed up. We did a big show for them. Some of the children were like "So what? So what?" They just… They take it for granted. They don't appreciate the effort gone in. And even that the extent of effort on the teacher's part to make enthusiastic sometimes isn't enough.
> …
> There's interaction, paired work and assessment. They were very focused. They talked to their partners. Once they got the function, they turned straight away and faced the teacher, and they're ready for the next bit. Not like in mine, when they finished talking, they'd talk about what was on telly last night.
>
> (EN-FG1, 22 May 2013)

All five teachers in EN-FG2 thought the Chinese lesson was very much driven by the teacher and totally teacher-led, with "no real independence at all" and no differentiation. Teacher EN3 said the teacher should do "prior-learning" to see whether the pupils knew it already, because he had a sense that they "kind of already knew it". Every time a teacher gave an opinion – most of the time a negative one, other teachers would immediately agreed to it, saying, "Yeah." The only "positive" point they seemed to all hold was about the discipline, but this was expressed in a tone of joking. When Teacher EN4 said she "like" the way the pupils all did exactly what they were told. Teacher EN3 added:

> The behaviour? That was impeccable, wasn't it? To answer a question, was literally I put up my hand and I stand up, which is really interesting to see.
>
> (EN-FG2, 6 June 2013)

Three other teachers immediately responded with a smile and "yeah". It is observable that English teachers in either group were inclined to agree with their colleagues and did not attempt to offer different views from others within a group. Two clear distinctions between two English focus groups were (a) EN-FG1 teachers were more willing to talk, whilst EN-FG2 teachers talked less on both focused lessons probably because of the presence of Teacher EN2 and (b) EN-FG1 teachers talked more about the positive aspects of the focused Chinese lesson whereas EN-FG2 teachers commented more on its negative aspects.

5.5 Chapter conclusion

In this chapter, we have heard a rich collection of voices regarding the effectiveness of mathematics teaching. These include the researcher's unstructured analysis of mathematics lessons, teachers' teaching beliefs and self-comments on their own lessons, their international awareness and attitudes to change and teachers' collective views on two lessons each from one of the two countries. Next chapter is the last chapter which will address the three research questions and conclude the study by highlighting the contributions of the study to research, practice and policy, clarifying its limitations and identifying directions for future research. Results and findings of the study will be synthesised with those of previous studies reviewed wherever relevant.

CHAPTER 6 - CONCLUSIONS

Chapter overview

6.1 Chapter introduction

This chapter consists of five sections. After this introductory section, Section 6.2 addresses the three research questions: the correlations between maths teaching & learning (RQ1), multiple perspectives on effective maths teaching (RQ2) and interconnections between evaluations and perspectives (RQ3). Section 6.3 focuses on the project's contributions to research on the effectiveness of mathematics teaching and its implications for practitioners and policy makers. Section 6.4 points out the limitations of the study and potential directions for future enquiries. Section 6.5 sees the thesis landing on its final remark.

6.2 Back to the research questions

The EMT project has collected data from 10 English teachers along with their 236 pupils and 9 Chinese teachers together with their 343 pupils across mathematics classrooms in two cities that were each socio-economically equivalently positioned in their home countries, England and China. It applies six methods – structured lesson observations (M1), questionnaires (M2), standardised mathematics tests (M3), unstructured lesson observations (M4), video-stimulated interviews (M5) and video-stimulated focus groups (M6) – to address the three major research questions and their corresponding sub-questions (the complete set of RQs may be found in Section 1.6):

- *RQ1: correlations between maths teaching and learning*
  - RQ1a: evaluating teacher behaviours
  - RQ1b: evaluating learning outcomes
  - RQ1c: correlating teaching with learning
- *RQ2: multiple perspectives on effective maths teaching*
  - RQ2a: the researcher's perspective
  - RQ2b: the teacher's perspective
  - RQ2c: native & foreign colleagues' perspectives
- *RQ3: interconnections between hard evaluations and soft voices*

The remaining part of Section 6.2 will draw evidence from the results and findings of the study to address the research questions sequentially.

6.2.1 Correlations between maths teaching & learning (RQ1)

This section has three subsections, each focusing on a sub-question of RQ1 respectively: the evaluation of teacher behaviours (RQ1a), the evaluation of learning outcomes (RQ1b) and correlating teaching with learning (RQ1c).

6.2.1.1 The evaluation of teacher behaviours (RQ1a)

The study evaluated the quantity and quality of teacher behaviours, as observed in nineteen maths lessons in Southampton and Nanjing, with two internationally validated observation systems – Opportunity to Learn (OTL) and International System for Teacher Observation and Feedback (ISTOF).

OTL results

OTL focuses on six percentages in relation to a lesson: the percentages of five types of classroom activities and the average percentage of pupils on task at regular intervals throughout the lesson. The five types of classroom activities are: *whole class interaction*, *whole class lecture*, *individual/group work*, *classroom management* and *partial class interaction*.

Pooling both countries' data together, the OTL evaluation found that English teachers allocated about a quarter (23.8%) of lesson time for whole class interactions and a half (46.6%) of the time for pupils working on their own without sufficient teaching input and that Chinese teachers spent about three quarters (72.2%) of lesson time interacting with the whole class, leaving about a quarter (27.8%) of the time for individual/group work. Moreover, English teachers spent a varying amount of lesson time lecturing to the whole class (3.8%), managing the class (3.5%) or interacting with individuals or part of the class (22.3%), whereas no Chinese teachers allocated any time for any of the three types of activities. Notably, the average time in which English teachers interacted with individuals or part of the class occupied approximately one fifth of the lesson time, and the time for whole-class interaction was almost the same amount. In addition, on average, the proportion of English pupils on task was 92.8%, whereas almost all Chinese pupils (99.7%) were on task. Since the class size in China was larger, the number of pupils off task per class was much fewer in China than in England. Only one English classroom had 100% pupil time on task, with the percentages for the rest of English classrooms ranging from 85% to 99%. Just one pupil in each of two Chinese classrooms was off task over a short period, and in the other seven classrooms all pupils were on task throughout the lesson.

ISTOF results

The revised version of ISTOF focuses on six dimensions of 40 effective teacher behaviours: assessment and evaluation (ISTOF 1), clarity of instruction (ISTOF2), instructional skills (ISTOF 3), promoting active learning and metacognitive skills (ISTOF4), classroom climate (ISTOF5) and classroom management (ISTOF6). All teachers' behaviours are each evaluated on a five-point scale, ranging from "strongly disagree = 1" to "strongly agree = 5" and clustered around 19 effectiveness indicators.

In ISTOF1, on average, English teachers score 12.8 (*SD* = 2.3), and Chinese teachers 19.4 (*SD* = 1); in ISTOF2, English teachers get 16.4 (*SD* = 4.4), and Chinese teachers 27.9 (*SD* = 1.7); in ISTOF3, the English mean is 11.4 (*SD* = 5.1), and the Chinese 23.2 (*SD* = 1.7); in ISTOF4, English teachers reaches 21.7 (*SD* = 6.9), and Chinese teachers 47.3 (*SD* = 2.9); in ISTOF5, English teachers average 22.8 (*SD* = 6), and Chinese teachers 38.8 (*SD* = 1.1); in ISTOF6, English teachers get 18.1 (*SD* = 6.4), and Chinese teachers 34.2 (*SD* = 1.7). Chinese teachers outperform their English colleagues in each of the six ISTOF domains. English teachers also have larger standard deviations than their Chinese colleagues, which suggests a larger within-country difference in England.

6.2.1.2 The evaluation of learning outcomes (RQ1b)

In order to assess 9 to 10-year-old children's maths performance internationally, a 40-item test was derived from the released item bank of TIMSS 2003 Grade-4 maths which was designed for children at the same age. The test was constructed by strictly following the proportions of content and cognitive domains defined by TIMSS. The initial intention was to assess both learning outcomes and learning gains with two tests scheduled at the beginning and end of ten school weeks in the second semester of the 2012-2013 school year. The results suggested a ceiling effect (Figure 4.38) upon Chinese pupils in Test 2 which prevented the actual measure of Chinese pupils' improvement over ten weeks. It is possibly because the items were too easy for Chinese children. The two test results were therefore correlated with teaching measures separately to evaluate teaching impact on learning outcomes at two points in time rather than gains over time.

The 40 TIMSS items included in the two EMT tests fall into five content domains – *Number* (n = 16), *Patterns and relationships* (n =6), *Measurement* (n = 8), *Geometry* (n = 6) and *Data* (n = 4) – and four cognitive domains – *Knowing facts and procedures* (n = 8), *Using concepts* (n = 7), *Solving routine problems* (n = 17) and *Reasoning* (n = 8). Correct rates (%) were calculated at the item, content domain, cognitive domain and paper levels. A pupil's paper-level correct rate stands for the overall test score for her/him.

At the item level

Each country's correct rate on every item is not only mutually compared but also compared with the TIMSS 2003 international average. In ***Test 1***, England has lower correct rates (*difference* ∈ [1%, 72%]) on 36 items than China and higher (*difference* ∈ [1%, 38%]) on 21 items than the TIMSS 2003 average; China has lower rates (*difference* ∈ [17%, 49%]) on 3 items than England, a same rate on item #17 as England's and lower rates (*difference* ∈ [10%, 11%]) on 2 items than the TIMSS 2003 average, exceeding the TIMSS average on 38 items (*difference* ∈ [7%, 67%]). In ***Test 2***, England's correct rates are lower (*difference* ∈ [4%, 62%]) on 35 items than China and higher (*difference* ∈ [1%, 45%]) on 31 items than the TIMSS 2003 average; China gets lower rates (*difference* ∈ [1%, 43%]) on 5 items than England and exceeds the TIMSS average on all 40 items (*difference* ∈ [2%, 75%]). In either test, China has a higher correct rate than England on 35 or more items.

At the domain level

Two countries' corrects rates in various domains are compared alongside the EMT international means. In ***Test 1***, England outperforms China in one of five content domains – *Data* – by 12% and has lower rates in the remaining four content domains (*difference* ∈ [23%, 36%]) and all four cognitive domains (*difference* ∈ [17%, 37%]) than China and in all content (*difference* ∈ [12%, 18%]) and cognitive (*difference* ∈ [9%, 19%]) domains than the EMT pooled average. In ***Test 2***, England still makes higher correct rates (*difference* = 12%) than China and than the EMT average (*difference* = 6%) in *Data* but lower in other content (*difference* ∈ [9%, 19%]) and cognitive domains (*difference* ∈ [13%, 30%]) than China and lower than the EMT average (content domain *difference* ∈ [10%, 14%], cognitive domain *difference* ∈ [6%, 15%]).

At the paper level

For every country, the mean of all papers' correct rates and the spread of children's performance are calculated. In Tests 1 and 2, English pupils have achieved a mean of 56 (SD = 21) and 66 (SD = 19) per cent respectively, and Chinese pupils averaged 83 (SD = 10) and 87 (SD = 10) per cent. In both tests, Chinese pupils outperformed their English peers by over 20 per cent with merely approximately a half of English standard deviation. Chinese pupils might have achieved an even higher mean in each test, if not because of the context-specific issues in items #9 and #10 (see section 4.5.1). The international differences on both occasions were statistically significant with a strong effect (Test 1: $p < .001$, $d = 1.74 > 1$; Test 2: $p < .001$, $d = 1.44 > 1$).

The overall results are in line with the identified gaps in mathematics performance between England and China in previous international studies (Lapointe et al., 1992; Lapointe et al., 1989; OECD, 2010, 2013).

6.2.1.3 Correlating teaching with learning (RQ1c)

To evaluate the effects of teaching on children's learning outcomes in mathematics, every teacher's OTL percentages and ISTOF domain scores are correlated with the class mean in each of the two maths tests cross-nationally.

Correlating OTL results with maths performance

Pooling data from two countries together, the correlational analysis indicated that two OTL components had a significantly positive correlation with pupil performance in either mathematics test, and these components were *whole-class interaction* (r = .97 or .91, p < .01) and *pupil time on task* (r = .95 or .91, p < .01). An OTL component – *whole-class lecture* (r = -.91 or -.92, p < .01) – posed a significantly negative effect upon mathematics scores in both tests, and two others – *individual/group work* (r = -.81, p < .05) and *classroom management* (r = -.77, p < .05) – were significantly related to lower scores in test one.

Correlating ISTOF results with maths performance

All six ISTOF components were found to correlate with higher maths scores in both tests. Pearson's r correlation coefficients of the six ISTOF components varied between .89 and .99 across two tests, the correlations between ISTOF measures and pupil maths scores were all significant at the level of .01. The greatest coefficients were of *promoting active learning and developing meta-cognitive skills* (r = .99 or .93) and *classroom climate* (r = .97 or .94).

6.2.2 Multiple perspectives on effective maths teaching (RQ2)

This section reflects upon various views regarding the effectiveness of maths teaching from the researcher, the teachers themselves and their local and international colleagues.

6.2.2.1 In the researcher's mind (RQ2a)

The researcher's unstructured observations focus on the flow and the characteristics of each lesson. In either aspect, there emerges a strong sense of national patterns.

Lesson flows

English lessons all started from lesson targets – what teachers called Success Criteria, went through short and speedy first input and then long repetitive re-inputs, and ended in children's self-assessment against "Success Criteria". Chinese lessons all started from real-life situations, went through intensive whole-class interaction around two to three exemplary tasks each followed by one or two similar tasks for independent work and ended in real-life situations. In the English class, children's independent work took a considerably longer period than in the Chinese class where the teaching input took relatively longer time and was carried out systematically and thoroughly so children could easily tackle similar problems quickly afterwards.

Lesson characteristics

In the *English maths class*, teachers (1) delivered differentiated teaching content, (2) shared all learning targets with the whole class at the very beginning of a lesson, (3) tended to transmit the knowledge directly to children, (4) re-taught individuals what had been just taught as many times as being asked for, (5) were facing children with weak foundation and poor readiness for the new content, (6) tended to care less about the quality and accuracy of their handwriting on the board while demonstrating procedures and/or solutions, (7) allowed pupils to move frequently and freely in the lesson particularly during independent work and (8) had at least one teacher assistant available for special educational needs (apart from classes EN3, EN4, and EN10).

In the *Chinese maths class*, teachers (1) made way for pupils to think actively and independently in finding solutions, (2) asked lots of questions to externalise children's thinking and reasoning of mathematical problems and of mathematical thinking and reasoning itself (metacognition), (3) were facing children with solid foundation and absolute readiness for the new content, (4) utilised pupils' classwork as teaching resources, (5) took seriously the timing of their writing on the board, i.e. when to write what, (6) spent most of the lesson time on teacher-guided and pupil-centred whole-class discussion, (7) facilitated less physical movement of pupils during the lesson and (8) had a large class size with little support for special educational needs.

6.2.2.2 In the teacher's mind (RQ2b)

Individual interviews with teachers collected teachers' perspectives in three main aspects: teacher beliefs in EMT, self-evaluation of the observed lesson and international awareness and attitude to change.

Teacher beliefs in EMT

Criteria of an effective maths lesson. English teachers thought an effective maths lesson should be relevant to real life, differentiated and child-centred, whilst Chinese teachers thought it should be aimed for and well accepted by pupils, be well planned, have real-life and long term meaning and have realised lesson targets.

Strategies of organising and managing lessons. English teachers all pointed out the strategy of assessment-aided differentiation, whereas Chinese teachers offered more approaches: paired discussion, the use of warm reminder, jokes or riddles, specific strategies for specific needs of each lesson, promoting peer instruction among pupils, and being well prepared and flexible.

Strategies for differentiation. English teachers talked about four ways of differentiation: self-differentiation, pre-assessment, support from the TA, and setting different tasks for different-ability children; Chinese teachers aimed for the majority, applied questioning techniques to offer opportunities for all-ability children, gave post-lesson support to low-ability and special awards to high-ability pupils, and might suggest low-ability pupils to preview the lesson content.

Teacher self-evaluation of the observed lesson

Lesson plan. Apart from Teachers EN1 and EN2 who both planned to teach times tables, all other English teachers did not have same planned lesson content. Chinese teachers were more likely to share the same teaching content in and across schools, probably due to the existence of Lesson Plan Groups within each school and the unified textbooks within each province.

The implementation of the lesson plan. Whilst all teachers thought their lesson plans were implemented well, many of them also reflected upon some flaws. English teachers were more likely to find their intended content did not match the levels of some or all children; Chinese teachers were more likely to find imperfection in small steps of the process which they thought as crucial for nurturing logical thinking. The central issue for English teachers seems to be *what* to teach and teach *whom what*, whilst that for Chinese teachers is likely to be *how* to teach. The former is rooted in the level of curriculum content and the corresponding differentiation, and the latter behaviours of teaching specific content. The former offers *certain* children opportunities to learn *certain* knowledge at *certain* levels, and the latter focuses on the *coherent* development of children's thinking in mathematics and offers *same* learning opportunities to *all*.

The effectiveness of the observed lesson. All teachers thought their lessons were effective, and all but two English teachers (EN1 & EN2) also spotted something imperfect. Again, as shown in their reflections upon their lesson plans, English teachers were more concerned with differentiation, whereas Chinese teachers were more worried about the function of specific teaching steps.

International awareness and attitude to change

The imagined foreign lesson. English teachers imagined that Chinese pupils would sit in rows either individually or in pairs, face the front, not be allowed to move and be well-motivated, on task and hard-working, that Chinese lessons would be very much teacher-led and structured and that there would be textbooks and more advanced content. Chinese teachers imagined that English pupils would be enjoying more freedom, sitting around tables in groups and receive plenty of individual care, that English lessons would be more relaxing and have looser standards and that the English class would be quite open and have a smaller class size.

Teacher flexibility to change. All teachers expressed their willingness to change, but, at the same time, two thirds of Chinese teachers also pointed out various things that might prevent them from thinking of change immediately and tremendously, such as the adaptability of successful approaches from other cultures in their own culture.

Overall, Chinese teachers tended to reflect on what had not been done well and know why it was that and how it might be improved immediately in the next lesson. English teachers were also concerned with certain aspects that were not ideal, but it seemed that those aspects, mostly extrinsic – such as differences among pupils' ability, had existed for long in their classrooms, hence seemingly impossible to change immediately.

6.2.2.3 In the local and foreign colleagues' mind (RQ2c)

By presenting two video-recorded lessons –one from England and the other from China – to teachers in four focus groups across the two geographical locations, local and foreign colleagues' views were collected regarding the effectiveness of the two lessons.

Collective views on the English lesson

Apart from Teacher EN2 and her colleague Teacher EN1, all other English teachers thought that this lesson was time-consuming and that teaching was almost absent. On the other hand, Chinese teachers had more diverse views on it. Out of nine Chinese teachers, three thought the lesson was ineffective and would lead to poor results; three regarded it as child-centred and individualised; two feared that this way of teaching would be unfair for low-ability children and would widen the performance gap among children over time; two saw the lesson as relaxing; one thought it could be effective if evaluated against the English standards; one thought this approach was rather time-consuming.

Collective views on the Chinese lesson

Half of English teachers thought the Chinese lesson was too much teacher-led and lacked differentiation; all of them thought the classroom discipline was impressive; the other half thought the Chinese children were self-motivated, eager to learn, and very focused and that the Chinese lesson was effective even though it had much fewer resources than the English – "it's very simple, but it's really effective." Chinese teachers talked a lot about details, interconnections between details, and the impact of the lesson upon children's development in and beyond the learning of mathematics. All agreed that the lesson would be effective in terms of correct rates, almost all pointed out both pros and cons of the lesson and reflected upon similar dilemmas they faced in practice, some gave suggestions about ways of optimising specific steps of the lesson, and some even scrutinised possible reasons for certain phenomena in the lesson.

Overall, English teachers' views were clustered within groups, whereas Chinese teachers' views were diverse and complementary to each other and their thinking were professionally deeper and more critical.

6.2.3 Connecting numbers with voices (RQ3)

This section addresses the third research question by interconnecting key results and findings from quantitative and qualitative parts of the study.

6.2.3.1 Differentiation & ability gaps (EN)

The fact that English teachers spent an average of 22.3% (*SD* = 25.4%) of lesson time on partial-class interaction fits in English teachers beliefs in the importance of teaching differentiation. In the English classroom, the constant streaming of children during the teaching process offered fewer opportunities for certain children to learn certain content, which over time might widen the performance gap. The much wider performance gap is evident in English children's test results, which is also in line with the ability picture that English teachers sketched in the teacher questionnaire on questions #35a and #35b (Table 4.21).

6.2.3.2 Pro direct transmission, re-inputs & partial interaction (EN)

English teachers were pro direct transmission, which was evident in the teacher questionnaire and unstructured lesson observations. Because they took the direct transmission approach, their cycle of delivering a set of knowledge was quick – too quick to get children's thinking thoroughly developed. The majority of children just didn't get it, and the speedy first input did not work, so the teacher had to re-do the teaching time and time again to individuals or part of the class in the same way. As the model of teaching did not fundamentally change, the following re-inputs did not seem to work well either.

6.2.3.3 Child-led or not? (EN)

The universally practised Success Criteria across the English classrooms echoes English teachers' strong belief in the child-led explorative approach. However, this belief contradicts with their beliefs in both the direct transmission of knowledge and the differentiation of teaching through individual interactions, in that the actual teaching does not allow children to lead the thinking process. In the

English classroom, it is the teacher that holds the standard answer/solution and demonstrates it to the class. Therefore, when all these beautiful ideas come together in practice, the intended learning simply does not happen.

6.2.3.4 Teaching for all, pro constructivist & whole class interaction (CN)

Chinese teachers' focus on all level children as expressed in interviews is consistent with the fact that they have spent an average of 72.2% (*SD* = 9.3%) of lesson time *interacting* with the whole class. Also, as found in the questionnaire data, they are pro constructivists, which can be triangulated by their teaching processes. The use of whole-class interaction and discussion offers more opportunities for them to externalise and develop children's mathematical thinking thoroughly on one task before moving on to the next. The intensive questioning and answering in the whole class enriches children's thinking and reasoning experiences, with them frequently explaining their ideas or listening to others' ideas, which allows children to mutually construct knowledge throughout the lesson. It reflects teachers' constructivist beliefs (as shown in Teacher Questionnaire) – that underpin their teaching processes – [TQ #29*l*] *thinking and reasoning processes are more important than specific curriculum content* and [TQ #29f] *pupils learn best by finding solutions to problems on their own* (Table 4.16 and Figure 4.25). The fact that the belief and practice of TQ #29*l* work is not because the curriculum content is unimportant, but because mathematics in itself is so highly coherent and logical that it is only through thinking and reasoning of it that children can get the content. During the thinking and reasoning processes in the whole class, the guidance of the teacher is the key to effectiveness and the *catalyser* to the development of children's thinking within limited lesson time. The Chinese way of teaching puts the mission of finding knowledge tightly in the children's hands instead of the teacher's. The Chinese belief and practice of letting children find solutions on their own (TQ #29f) is neither through children's independent work nor through the teacher's direct delivery of intended content. Rather, it is through the teacher-guided child-centred approach[11] in the course of whole class interactions.

6.2.3.5 The effectiveness of first input & teachers' subject matter knowledge

English teachers spent less time on the first input of knowledge, which could be partly explained by their lack of sufficient subject and pedagogical content knowledge. The questionnaire data also confirmed that English teachers were not specialist maths teachers and that they, however, expressed a lower level of needs in developing their subject matter and pedagogical content knowledge than did Chinese teachers who were all specialist teachers. Chinese teachers all spent longer time on the first occasion of knowledge input, which was not only observable in all lessons but also interpretable in teacher interviews. For instance, Teacher CN1 said, "I must make sure every pupil clearly understand the method through the teaching and learning of ***one*** example …" In the centre of a Chinese lesson were intensive whole-class interactions and discussions around no more than four examples which were in a hierarchical order and did not simply repeat

[11] this approach is what Chinese teachers commonly call "以教师为主导，学生为主体".

each other. This is consistent with the finding that more effective teachers have limited content focus per lesson in the Junior School project (Mortimore et al., 1988). Chinese teachers ensured that children's *first* opportunity to learn must be successful, which echoes with similar findings of high success rates in early American TER studies (Berliner et al., 1978; Fisher et al., 1978) and with similar principles held by Chinese maths teachers in Ma's work (1999, p.46) who "pay particular attention to the first time an idea is introduced to students in its simplest form".

On the contrary, English teachers' first input did not result in pupils' mastery of the delivered knowledge. Although they thereafter gave children multiple opportunities to learn the same content, the ineffectiveness seemed always there – as if a shadow to the unchanging teaching processes. Because of English teachers' lack of subject matter and pedagogical content knowledge, the first input was not well structured, the big aim was not broken into digestible smaller chunks, the lesson components did not interconnect mutually coherently, the time applied on each occasion of input was too short for children to think thoroughly, and the questions from the teacher lacked necessary depth. Such a connection between profound subject knowledge and well-structured lesson was also found in previous studies (e.g., Ma, 1999; Tikunoff et al., 1975).

6.2.3.6 **Lesson objectives, direct transmission & constructivist**

The way two countries' teachers present lesson objectives also reflect their beliefs. All English teachers presented directly and clearly, at the very beginning of each lesson, the content and targets that would be covered and were expected to achieve. This falls in the format of direct transmission. Conversely, most Chinese teachers did not clarify the lesson targets at the beginning. Rather, they started from Q & A about real-life issues or phenomena and pointed out the title of the lesson when relevant answers emerged, the detailed knowledge points would appear one by one on either a slide or the chalkboard as they taught, and the lesson objectives gradually entered the scene one at a time as the lesson went on.

6.2.3.7 **Self-reflection & metacognition**

Both countries' teachers facilitated pupils' self-reflection upon learning, but there were considerable differences. English teachers put more emphasis on reflecting upon the lesson goals and results – what I'll achieve today (at the beginning), what I have achieved today, and what I aim to achieve tomorrow (at the end). Nonetheless, Chinese teachers stressed more on the process and reasons behind it – how you have done it, how someone has made that mistake and why, what is the reason for learning this, and so forth. English teachers often asked pupils to make reflections at the beginning and the end of the lesson, while Chinese teachers kept promoting pupils to do so throughout the lesson. It is thus not surprising that Chinese teachers scored higher in ISTOF in promoting metacognitive skills, because thinking of thinking processes – rather than thinking of learning outcomes (the *right* facts) – is the key to developing metacognition.

6.2.3.8 **Teacher questioning: who solve the problem?**

The most significant distinction between the English and the Chinese approaches perhaps lies in the way of teacher questioning. Typical scenarios would be that the former threw a question to one or several individual(s) and soon chose to explain the answer or procedure her/himself if the ‘correct’ answer did not emerge, whereas the latter not only asked more questions but also constantly promoted the children to find the answer(s)/solution(s) themselves through intensive questioning and answering. In the English classroom, there is a standard procedure or solution that the teacher feels must be transmitted to the children; in the Chinese classroom, as aforementioned, there are multiple ways of solving a problem which are expected to emerge in the lesson *from* pupils. Chinese teachers were very *unlikely* to tell the pupils the “right” solution to *any* problem – big or small, even though the solution was the core target of the lesson. Chinese pupils generated and explained each solution during the process of teacher-initiated Q & A. The purpose of such intensive whole-class discussion and exploration was to get pupils to arrive at the lesson destination(s) by themselves. English teachers generally held the solution and passed it on to pupils directly while interacting with them. If all pupils could apply the solutions to similar problems, then the English lesson was deemed successful.

6.2.3.9 **Who’s in charge of what?**

English teachers were in charge of what should be the correct solution/procedure, though they let children be in control of the level of work they would like to work on. Whilst Chinese teachers were more likely to be in charge of the lesson rhythm/process, they let the children lead the way and think, find and tell every possible solution/procedure to every problem. In English maths lessons, there seemed to be one possible solution to a given problem; in Chinese maths lessons, there were many possible solutions to a single problem. In English maths lessons, the teacher held the unique key to a specific unknown and passed it on to children so that they could also unlock it in the same way; in Chinese maths lessons, the children held the key to trying multiple possible keys to an unknown with varying forms in varying situations. In the English maths lessons, the teacher played the main role of mathematical thinking and reasoning; in the Chinese maths lessons, it was the pupils that played this role.

6.2.3.10 **Who repeats what?**

English teachers tended to re-teach multiple times in a lesson the same content after the first input, whereas Chinese teachers were more likely to ask pupils to find the solution(s) and then re-explain it/them at different points in the lesson.

6.2.3.11 **Where to conclude?**

English lessons all had a final main conclusion in the end, while Chinese lessons all had a mini-conclusion near the end of each discussion throughout the lesson process and might or might not have a final conclusion at the end of the lesson.

6.2.3.12 **Classroom management & corresponding CPD needs**

English teachers on average spent 3.5% of lesson time managing the class but had lower percentages of pupils on task; Chinese teachers did not spend any time on classroom management but had higher percentages of pupils on task. Combining findings in observations and interviews, one could easily see that Chinese teachers were more capable of managing the class without apparent managing behaviours than English teachers. Nonetheless, in the questionnaire, Chinese teachers demanded more needs in developing their classroom management skills than did English teachers. It may be because they appreciate the importance of management skills and have been trying hard to improve such skills that they actually turn out to be better at it, which is similar to the case of subject matter and pedagogical content knowledge.

The differences in the amount and content of CPD might indirectly influence how teachers teach, and this is probably all CPD programmes intend to do. However, the maths class is the arena where formal teaching and learning of mathematics happens, and children learn maths from teachers' teaching directly rather than their CPD. To improve maths teaching, designers of primary maths CPD programmes may consider reinforce the effective teaching factors identified in the EMT study. School leaders should create an environment for more school-based CPD to happen and nurture a positive climate in which teachers are happy to get involved and continue to grow.

6.2.3.13 **Cultural explanations**

Researchers from the TIMSS Video Study come up with the idea that teaching is a *cultural* activity (Stigler & Hiebert, 1999). The EMT study proves that they are both right and wrong. They are right because the process of mathematics teaching varies considerably from England to China. They are wrong because they neglect the fact that culture is a collective habit of thinking and doing things and thus fail to conclude such differences at both macro and micro levels. At the macro level, it does seem to suggest that teachers in different cultures teach mathematics in different ways. However, at the micro level, the end products of a culture in the classroom are various elements of teaching and learning. Moreover, culture is not static; it is dynamic and constantly evolving. Members of a certain culture are influenced by the culture they perceive and believe in and at the same time are also constantly reformatting the culture with or without clearly defined purposes. Certainly, schools and classrooms also have their own mini-cultures, and the change of institutional cultures has increasingly been attempted by teachers and school leaders (Fullan, 2007). Cross-cultural studies are thus eye-opening, because they offer alternative options from other cultures and generate new ideas and methods in the process of observations and reflections.

6.2.3.14 **Homework differences**

The results of the pupil questionnaire show that children from two cities have different amount of homework, but research suggests that the amount of homework has approximately zero impact on pupils' achievement in primary schools (Cooper, 1989). Through meta-analyses, Cooper (1989) identified differentiated effects of the *amount* of homework on pupils' achievement at different schooling stages: (1) $r \approx 0$ for Grades 3 through 5 (Years 4 through 6 in England), (2) $r = .07$ for Grades 5 through 9 (Years 6 through 10 in England) and (3) $r = .25$ for upper secondary school. It was recommended that the amount of weekly homework should be (i) one to three assignments (each lasting no more than 15 minutes) for lower primary pupils, (ii) two to four assignments (each lasting 15 to 45 minutes) for upper primary pupils, (iii) three to five assignments (each lasting 45 to 75 minutes) for lower secondary pupils, and (iv) four to five assignments (each lasting 75 to 120 minutes) per week for upper secondary pupils.

Though studies do suggest a positive effect of homework on achievement (Cooper et al., 2006; Good & Grouws, 1979), the positive effect will not happen unless the homework is of an optimal amount and good quality and has reasonable purposes. Otherwise, the effect can be negative (Hallam, 2004). Comparing with time on task and other factors, the effect size of homework also tends to be smaller (*ibid*. as cited in Muijs & Reynolds, 2011, p. 142).

In summary, the amount of homework will have little impact on children's learning if they have not grasped thoroughly the knowledge underpinning their homework. In the EMT study, children in the English classroom kept looking for re-inputs from the teacher, which had already alerted the ineffectiveness of teaching. The amount of homework would mean nothing to lots of (in some classes almost all) English pupils after class. In the Chinese classroom, children spent the majority of lesson time gradually understanding the knowledge through deeper thinking and actively answering inter-connected questions; their seatwork was quickly and evenly checked by the teacher; work samples with right or wrong solutions were picked up, presented and discussed in the whole class. After class, the Chinese pupils demonstrated a thorough understanding of the knowledge taught and they were ready for independent homework regardless the amount. Without the mastery of knowledge in the class, the amount of homework would add little positive value to pupils' maths performance.

6.2.3.15 **Summary**

Overall, English teachers tended to emphasise the importance of differentiation and the idea of child-centred teaching, however they seemed to have over contradicted whole-class interactive teaching with the child-centred idea. As a result, they seemed to uniformly believe that putting children at the centre of the class could only be realised through individualised interactions. Chinese teachers on the other hand were more concerned with detailed steps of teaching and saw the steps of teaching as a coherent, irreversible and logical whole, just like the hierarchical nature of mathematics itself. They also put the children at the centre of the teaching and learning, but their emphasis was on the children taking the lead of the problem-solving process in the scope of the whole class. The teacher-pupil and pupil-pupil interactions in the whole class demanded everyone's full attention, as the questioning and answering was both brisk and engaging. Their approach is more of scaffolding and constructivism than direct transmission, whereas the English approach is more of direct transmission – which is surprising but is evident across the ten English classrooms. English teachers see the differences among pupils as an obstacle and therefore feel the necessity of teaching them separately according to their levels; Chinese teachers see the differences among pupils as natural resources of teaching and learning, that contribute to the progress of everyone in the class, and thus are quite comfortable about putting children together and interacting with them actively.

6.3 Contributions & implications

The study drew on both quantitative and qualitative strands of data collected in mathematics classrooms for 9- and 10-year-olds in two socio-economically equivalent cities from two cultures and two countries. It correlated teacher behaviours with mathematics learning outcomes, captured multiple perspectives from different roles regarding the effectiveness of mathematics teaching within and across national borders and scrutinised the interconnections between hard measurement and soft voices. This section summarises the *contributions* of the EMT project to research in the effectiveness of mathematics teaching and its *implications* for practitioners and policy makers on how to improve mathematics teaching and learning.

6.3.1 Contributions to research

The study advances the fields of teaching effectiveness research and mathematics pedagogical research in but not limited to the following eight aspects:

- Collecting cross-national evidence on effective maths teaching in primary schools
- Replicating effective teaching 'elements' found in previous national studies,
- Discovering TER evidence from the West in an Eastern context
- First TER study attempting the value-added approach across nations
- Applying both teaching effectiveness and improvement methods
- First TER study seeking to form an international dialogue on EMT
- Reuniting two communities of maths teaching research
- Methodological contribution: multifunction of videos in TER

6.3.1.1 Cross-national evidence

The study filled in the gap in the literature where international comparisons lacked sufficient attempts to interconnect teaching and learning in one study, where the evidence of TER studies was mainly drawn from national studies and barely crossed nation borders, and where hard evaluations of correlations between classroom variables lacked the crucial support from soft but rich explanations that only qualitative data may offer. It is the first study after ISERP (Reynolds et al., 2002a) that has ever collected TER data cross-nationally and brings to light in time cross-national empirical evidence that the world is longing to know regarding what works in the teaching of mathematics internationally. It is also the first TER study that has ever attempted to evaluate the value that different teaching approaches might add to pupils' mathematics performance *cross-nationally*.

6.3.1.2 **Prior evidence replicated**

The results and findings of the study further confirm what have been repetitively found as effective at the classroom level in previous studies conducted within different countries. It reassures that certain teaching factors do travel internationally as foreseen by the ISERP research team (Reynolds et al., 2002a). It consolidates the foundation of TER and the development of educational science in the subject area of mathematics cross-nationally. It tests the adaptability of two established observational instruments – OTL and ISTOF – in two geographically and culturally considerably different countries. The pooled analyses of two countries' data coincide with both the underlying hypotheses of the two instruments and the existing evidence in the literature.

Trace the OTL correlates back to the literature. High opportunity to learn has already been found in more effective teachers' classrooms about four decades ago in that it guarantees both quantity and quality of children's engagement into the academic content that they are learning (Arehart, 1979; Fisher et al., 1980; Stallings et al., 1977). The study by Good and Grouws (1979) indicated that whole class instruction could only be effective if the teacher taught in an *interactive* way. This is evident in this study's OTL correlates in that whole class *interactive* teaching predicted higher maths performance whereas whole class *lecture* led to lower maths performance. The British study, MEPP by Muijs and Reynolds (2003), also found the positive effect of whole class interactive teaching upon both learning outcomes and gains, which is consistent with partial findings in previous TER studies in the UK and US (Croll, 1996; Emmer et al., 1979; Galton et al., 1980; Pollard et al., 1994). The literature also suggests that teachers who spend more time interacting with the whole class tend to have higher rates of time on task (Croll, 1996; Muijs & Reynolds, 2003; Pollard et al., 1994). Conversely, teacher interaction with individual pupils poses a negative effect upon *time on task* (Croll & Moses, 1985) which is however crucial in nurturing pupils' thorough engagement into subject learning rather than social activities and promoting optimal learning outcomes/gains (Muijs et al., 2014; Muijs & Reynolds, 2003).

Trace the ISTOF correlates back to the literature. More effective teachers are good at asking questions to assess the status of pupils' understanding and guide them to continue thinking and reasoning through follow-up questioning, which is also evident in the literature (Evertson et al., 1980; Muijs et al., 2014). A lesson with a clear and logical structure follows the way the human brain processes information, hence contributing to the quality of teaching (Clarke et al., 1979; Land, 1979). Effective instruction happens in the classroom where the teacher engages and interacts with *all* pupils through appropriate questioning and answering and thus develops their understanding of intended academic knowledge (Fisher et al., 1980; Stallings & Kaskowitz, 1974). The effective way of promoting active learning and metacognitive skills is also mainly through teacher questioning. Effective teachers tend to ask more process questions which lead children to more active reasoning and critical thinking (Evertson et al., 1980; Muijs et al., 2014; Muijs & Reynolds, 2000). More effective teachers are good at building a learning-oriented climate through consistently effective management of time and pupils' learning momentum and steadily effective interactions with all pupils (Brophy, 1973; Muijs et al., 2014; Muijs & Reynolds, 2000). More effective teachers are good at managing time, classroom activities and pupil behaviours (Berliner & Tikunoff, 1977; Good & Grouws, 1977); the EMT study further suggests that they are capable of preventing off-task misbehaviours from happening through highly interactive teaching with the whole class, and thus no apparent managing actions are needed.

6.3.1.3 The discovery of TER evidence from the West in an Eastern context

The majority of effective teaching elements have been long discovered in the West using the teaching effectiveness research approach. They are for the first time discovered in China – a context that is generally believed to be quite different from the West. This reassures that certain teaching factors do travel across nations and cultures.

6.3.1.4 First TER study ever attempting the value-added approach across nations

The study employs pre- and post-tests with a genuine intention to assess the academic value that teachers might have added to pupils' learning in maths over time (ten school weeks). It is also the first study that has ever conducted cross-study comparisons of English children's maths performance using exactly the same items – 40 TIMSS 2003 Maths items, which offers a historical view of the performance of 9 to 10-year-olds in both content and cognitive dimensions over the course of nine years.

6.3.1.5 Advancing the TER field by applying effectiveness and improvement methods

Inheriting the traditional quantitative traits of TER, the study develops the field through incorporating both the teaching effectiveness (QUAN) and teaching improvement (QUAL) approaches. The study sees both approaches as parts of a whole rather than two mutually contradicted sides.

6.3.1.6 First TER study seeking to form an international dialogue on teaching

Across the nation border, the study has listened to multiple voices of individuals and groups of teachers about the same topic – what works in teaching maths in primary schools. This is the first TER study that has ever attempted to collect and appreciate practitioners' views across nations and cultures. About maths teaching in general and the effectiveness of two focused lessons in particular, teachers' voices have travelled across space and time and joined each other in this international dialogue in the EMT study.

6.3.1.7 Reunion of two communities of maths teaching research

This is not only one of the few TER studies that have attempted to understand the ways in which teachers handle the subject matter but also one of the few maths pedagogic studies that have systematically looked into what teachers *do* in the maths class. The EMT study interweaves all strands of data and evidence around maths teaching and appreciates both the content and the process of teaching. When it comes to maths teaching and learning, there are no either-ors – there have to be both. The study is another starting point for maths education researchers and effective maths teaching researchers to reconcile and work together for a better maths education for every child.

6.3.1.8 Methodological contribution: multifunction of videos in TER

The study demonstrates an optimal utilisation of video in classroom observations, post-lesson interviews and focus groups. The use of video might not occur to be foreign for researchers in TER and other communities in educational research. However, the use of video for teacher reflections and collective comments appear to be rare in the current TER community, partially because of its traditional interest in quantitative data. In the EMT study, video-recorded lessons have made the international dialogue on the effectiveness of maths teaching easily come true, with teachers sitting in their comfort of local schools. Moreover, video data have provided the study an opportunity to integrate different types and strands of evidence. The study has thus set an encouraging example for future TER studies.

6.3.2 Implications for practice & policy

The performance gap in mathematics between the West and East Asia has been revealed repetitively in various international surveys of student achievement over the past couple of decades. Building on data collected from two cities that were each positioned in a socio-economically equivalent place in their home countries, this study further confirmed the existence of such a gap between English and Chinese 9 and 10 year-olds. Policy makers from the West are increasingly concerned with the results of international assessments and have been seeking ways of improving educational outcomes nationwide. However, most innovative plans tend to stop at the national or school level, hence missing the key lever to change – teaching at the classroom level.

In England, practitioners and policy makers have been heavily influenced by the advocates for progressive and individualised teaching (Central Advisory Council for Education, 1967), a broader curriculum, the superficial likes and dislikes of international comparisons (Alexander, 2012) and the one-sided assertion that pedagogy is merely an individualised and cultural thing (Alexander, 2008). The curriculum needs to be broad, which, nevertheless, cannot be an excuse for the lack of depth and effectiveness of curriculum implementation. The intended content of every subject, including mathematics, should be thoroughly and firmly grasped by children after all efforts have been put in at the system, school and classroom levels.

The danger is not what Pasi Sahlberg superficially calls "GERM" (as applauded by Alexander, 2012). The real danger is the lack of standards on education and its evaluation and the unawareness or ignorance of ineffective educational processes. There will not be any improvement if educational practitioners, researchers and policy makers are ready to play ostrich whenever encountering disappointing outcomes.

It is fine to maintain a broad curriculum (reasonably broad, not simply broad***er***, time is limited). Nonetheless, England should, at the same time, ensure the high quality of teaching and learning of the core subjects – such as English and Mathematics – which form the base of many other subjects and children's more fulfilling future life in a more advanced world. When children are underperforming in these core subjects, it is time for all educational stakeholders to reflect, research and react on the effectiveness and the improvement of teaching and learning in these subjects. It is as simple as that. To solve the issue, one needs to look at and focus on the issue itself.

Empirical studies in the literature and the EMT study all suggest that the English pedagogy is severely individualised, despite a small reduction over an unbelievably long time span. The proportion of time in which children work on their own has dropped from 2/3 of the lesson time in the ORACLE study (Galton et al., 1980) to 46.6% in the EMT study. The amount of time that English teachers spend interacting with individuals has decreased considerably from 67% in the Junior School Project (Mortimore et al., 1988) to 22.3% in this study. In addition to the historical and contemporary advocates for progressive and individualised teaching, another reason for the heavy use of individualised activities in the English classroom might be that English teachers' lack of profound subject knowledge limits their confidence in interacting with the whole class and restrains their ability in breaking the intended main topic into small, digestible and hierarchical (as mathematics itself) chunks and organising questioning and reasoning activities in a coherent manner, with the whole class involved and engaged.

In fact, whether to interact with individuals or the whole class is ultimately not an either-or question when it comes to teaching and learning in the classroom. Each individual child matters – there is no doubt. However, before we could afford (even if we could, do we really have to?) to allocate a subject expert teacher for every single child and also before we could all decide to ignore the very fact that human beings are a social species who grow and evolve in social dialogues and activities not only as individuals but also as part of a whole, a considerable amount of lesson time should be allocated for the teacher *interacting* (**not** lecturing) with the whole class, if we are looking for the maximum progress of *every* individual child. In the dynamic of whole class interactive teaching and learning, individuals are open to all ideas and solutions that could possibly be generated by the whole class, hence progressing more rapidly and holistically.

Findings also suggest the difference of the subject matter and pedagogical content knowledge between two countries' maths teachers. English teachers are generalists and their subject matter and pedagogical content knowledge is weaker than their Chinese counterparts who are mathematics specialists. This is evident in the way they deal with the content in the classroom. There is a necessity for English primary schools to arrange specialist teachers across subjects so that teachers are more capable of optimising the quality and widening the scope of their interactions with the class. Specialist teachers will also have more time and space to prepare and reflect upon lessons in their specialised subjects, and children will benefit the most from what every teacher is most capable of.

It is part of the research rationale of the study to collect internationally applicable effective teaching behaviours in mathematics classrooms. The pooled analysis of lesson data from both countries means that those effective teaching approaches work both within and across countries. Those approaches are worth trying out (or keeping for the case of China) in either country's mathematics classrooms.

In England, effort needs to be put into increasing the proportion of time on whole-class interactive teaching, reducing that on individual work or teacher interactions with individuals and avoiding whole-class lectures, so that every individual can benefit the most from every lesson. For English teachers and policy makers, the big challenge/dilemma lies in two aspects: (1) whole-class interactive vs partial-class interactive or individual work and (2) specialist vs generalist teachers in primary school.

To get out of the dilemma, England needs to make choices. Recommendations based on decades of the TER literature and the EMT project findings are given here for English practitioners and policy makers to consider:

1) Avoid the utilisation of whole-class lectures;
2) Increase the proportion of lesson time on whole-class interactions;
3) Reduce the proportion of lesson time on individual/group work;
4) Reduce the proportion of lesson time on partial-class interactions;
5) Enhance teachers' subject matter and pedagogical content knowledge;
6) If possible, transform primary teachers into specialist teachers;
7) Make change in initial teacher education accordingly.

In China, attention should be placed on preserving the current teaching approaches in an increasingly Westernised (Americanised) trend in the country's education and society, increasing support to children with special education needs and preventing radical change from happening in the classroom. The problem in China does not lie in classroom teaching and learning. It lies in the political, social and cultural domains of the country.

6.4 Limitations & future directions

This section reflects upon the ***limitations*** of the project and point out possible ***directions*** for research in this field in the foreseeable future.

6.4.1 Limitations

The main limitations of the study lie in its sample size and statistical methods. It would have come up with richer findings if more schools from more regions in both countries had participated. The correlational method did not take other factors into consideration, such as the possible indirect impact of teacher professional development and/or teachers' subject knowledge on teaching and learning, the effect of homework marking and feedback on learning. Contextual variables, such as the degree of parental involvement, have not been included. Thus, the strength of a correlational study is its focus on key causes predicting specific results, which also becomes its weakness in that other essential factors and indirect predictors may have as well been ruled out.

The test instrument derived from TIMSS 2003 turned out to be too easy for the Chinese pupils. Subsequently, a ceiling effect has prevented the study from measuring learning gains. With a better set of test items, the study may have come up with more robust evidence for maths teaching improvement across nations in the long run.

Moreover, due to time limit and other conditions that might have limited a PhD study, only one lesson in each classroom was observed. Although the teachers were asked to deliver their lessons in their usual ways, a sequence of lessons from each class would have in themselves generated the typicality of the teacher's maths teaching.

In addition, the content of the observed maths lessons was not exactly the same, due to the difficulty of organising uniform teaching content across schools and nations. It would have generated deeper findings if all lessons were about the same subject area(s). Similarly, with the lesson content controlled as the same across classrooms and the test items designed for the same content as well, the study would have come up with effective teaching factors for specific maths topics for children at the specific age.

Last but not least, though every effort has been made, the study is still limited by the fact that only one PhD researcher – despite having the advice from the Supervisor – was independently carrying out this cross-national investigation. A team of researchers would have been able to generate more meanings from the multiple types of data. Though the researcher has made every effort to maintain a neutral position in collecting and interpreting the data, an international team would be much more helpful and efficient in working against potential biases.

6.4.2 Future directions

Future TER studies are demanded with larger samples of participants recruited, more countries involved and multi-level modelling methods applied. Teaching improvement research projects are also much needed to bring TER evidence into practice, re-test teaching effects on each sample, and bring educational researchers, teachers and school leaders together to make evidence-informed innovations in classrooms. Researchers may consider experimental designs that involve pre- and post-lesson, pre- and post-unit, pre- and post-term assessment to compare the effects of different teaching approaches with the control of teacher-level variables, such as subject matter knowledge and length of teaching experience, and pupil-level variables, such as prior performance and family background.

Classroom-level factors are largely not investigated across the East and the West. The reason that international studies across different cultures and countries are meaningful and worth conducting is different nations have different educational policies and realities which could be seen as naturally existing laboratories where different educational experiments are constantly carried out (IEA, 2012). Different experimental inputs lead to different outcomes (ibid.). Studies across a variety of countries offer more opportunities to find which experiment works while which doesn't and consolidate the knowledge base of educational science than studies within and/or across culturally similar countries. Given the typical performance gap between East Asian countries and Western countries, systematic large-scale TER studies will make meaningful contribution to the world's understanding of how to teach maths more effectively and scientifically.

This study only looked at the classroom level, so inter-country comparisons were only based on classroom level data. There are more to be explored in future to see correlations between independent and dependent variables at pupil, classroom, school, and system levels and between independent variables within and across these levels.

6.5 Concluding remark

The replicated findings about what works in the mathematics classroom make a wake up call to practitioners, policy makers, educational improvement researchers and all other educational stakeholders in both countries and beyond, regarding the possibility of making change in the teaching and learning of mathematics. The multiple perspectives among teachers indicates the connections between teachers' beliefs and their behaviours, which offer possible indirect ways of improving teaching and learning, for instance, encouraging teachers to enhance their subject matter and pedagogical content knowledge, come out of their comfort zones and try out new approaches. There is also a need to increase collaboration between researchers and practitioners for three reasons: (1) this will promote the circulation of research evidence between researchers and teachers, (2) this may nurture critical thinking among teachers and (3) researchers will have updated information of what is going on in the classroom 'right now', therefore being more capable of identifying and tackling key research issues. Crucial things that policy-makers can do are to look at both initial teacher education and continuous professional development of in-service teachers, to implement teaching innovations based on existent empirical evidence and to encourage and support more classroom studies for them to consistently evaluate what works, how and why.

REFERENCES

Alexander, R. (2008). Culture, dialogue and learning: notes on an emerging pedagogy. In N. Mercer & S. Hodgkinson (Eds.), *Exploring Talk in School* (pp. 93-114). London: Sage.

Alexander, R. (2012). Visions of education, roads to reform: the global educational race and the Cambridge Primary Review. Retrieved from: http://www.robinalexander.org.uk/wp-content/uploads/2012/04/Alexander-Chile-GMU-lecture-1-121030.pdf

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*(4), 193-223.

Arehart, J. (1979). Student opportunity to learn related to student achievement of objectives in a probability unit. *Journal of Educational Research, 72*(5), 253-269.

Askew, M., Rhodes, V., Brown, M., Wiliam, D., & Johnson, D. (1997). *Effective teachers of numeracy: Report of a study carried out for the Teacher Training Agency*. London: King's College, University of London.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In W. S. Kenneth & S. Janet Taylor (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89-195). London: Academic Press.

Becker, I. L. (1978). *Instructional ingredients for the development of beginning reading competence*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Becker, W. C. (1977). Teaching reading and language to the disadvantaged - What we have learned from field research. *Harvard educational review, 47*(4), 518-543.

Bell, J. (2010). *Doing your research project: a guide for first-time researchers in education, health and social science*. (5th ed.). Maidenhead: Open University Press.

Bennett, N. (1987). Changing Perspectives on Teaching□learning Processes in the Post□Plowden Era. *Oxford Review of Education, 13*(1), 67-79. doi:10.1080/0305498870130107

Berliner, D., Fisher, C., Filby, N., & Marliave, R. (1978). *Executive summary of Beginning Teacher Evaluation Study*. San Francisco: Far West Laboratory.

Berliner, D., & Tikunoff, W. (1976). The California Beginning Teacher Evaluation Study: overview of the ethnographic study. *Journal of Teacher Education, 27*(1), 24-30.

Berliner, D., & Tikunoff, W. (1977). Ethnography in the classroom. In G. Borich & K. Fenton (Eds.), *The Appraisal of Teaching: Concepts and process* (pp. 280-290). Reading, MA: Addison-Wesley.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook* of mixed methods in social & behavioral research (2nd ed., pp. 95-118). Los Angeles: SAGE.

Bill & Melinda Gates Foundation. (2012a). Asking students about teaching: student perception surveys and their implementation. Retrieved from: http://www.metproject.org/downloads/Asking_Students_Practitioner_Brief.pdf

Bill & Melinda Gates Foundation. (2012b). Learning about Teaching: Initial findings from the Measures of Effective Teaching project. Retrieved from: http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Bill & Melinda Gates Foundation. (2012c). Measures of effective teaching. Retrieved from: http://www.metproject.org/index.php

Bill & Melinda Gates Foundation. (2012d). Student Perception Survey for Elementary Students. Retrieved from: http://www.metproject.org/resources.php

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' Experiences of Ability Grouping - disaffection, polarisation and the construction of failure. *British Educational Research Journal, 26*(5), 631-648. doi:10.1080/713651583

Bogdan, R. G., & Biklen, S. K. (1992). *Qualitative research for education*. (2nd ed.). Boston, MA: Allyn & Bacon.

Boonen, T., Van Damme, J., & Onghena, P. (2013). Teacher effects on student achievement in first grade: which aspects matter most? *School*

Effectiveness and School Improvement, 25(1), 126-152. doi:10.1080/09243453.2013.778297

Borg, W. (1980). Time and school learning. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 33-72). Washinton, DC: National Institute of Education.

Bosker, R., & van der Velden, R. K. W. (1989). *The effects of schools on the educational career of disadvantaged pupils*. Paper presented at the The 2nd International Congress for School Effectiveness and School Improvement, Rotterdam, NL.

Boyer, C. (1968). *A history of mathematics*. New York: John Wiley & Sons.

Brewer, J., & Hunter, A. (1989). *Multimethod research: a synthesis of styles*. Newbury Park, CA: SAGE.

Brophy, J. E. (1973). Stability of teacher effectiveness. *American educational research journal, 10*(3), 245-252.

Brophy, J. E., & Evertson, C. (1976). *Learning from teaching: a developmental perspective*. Boston: Allyn and Bacon.

Brophy, J. E., & Good, T. L. (1986). Teacher behaviour and student achievement. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 328-375). New York: Macmillan.

Brown, A. L. (1994). The advancement of learning. *Educational Researcher, 23*(8), 4-12.

Bryman, A. (2006). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology, Theory and Practice, 9*(2), 111-126.

Cai, J. (2003). Investigating parental roles in students' learning of mathematics from a cross-national perspective. *Mathematics Education Research Journal, 15*(2), 87-106. doi:10.1007/bf03217372

Cai, J., Ding, M., & Wang, T. (2014). How do exemplary Chinese and U.S. mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265-280. doi:10.1007/s10649-013-9513-3

Cai, J., Moyer, J. C., & Wang, N. (1999). Parental roles in students' learning of mathematics: an exploratory study. *Research in Middle Level Education Quarterly, 22*, 1-18.

Cao, Z., Bishop, A., & Forgasz, H. (2007). Perceived parental influence on mathematics learning: a comparison among students in China and Australia. *Education Studies in Mathematics, 64*(1), 85-106.

Central Advisory Council for Education. (1967). *Children and their primary schools: a report*. London: H.M.S.O.

Centre for Cities. (2015). Southampton Factsheet. Retrieved from: http://www.centreforcities.org/city/southampton/

Chen, C., & Stevenson, H. (1995). Motivation and mathematics achievement: a comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development, 66*(4), 1215–1234.

Chen, X., & Li, Y. (2010). Instructional Coherence in Chinese Mathematics Classroom - A Case Study of Lessons on Fraction Division. *International Journal of Science and Mathematics Education, 8*(4), 711-735. doi:10.1007/s10763-009-9182-y

China Knowledge. (2014). Nanjing City Information. Retrieved from: http://www.chinaknowledge.com/CityInfo/City.aspx?Region=Coastal&City=Nanjing

Clark, R. E. (1982). Antagonism Between Achievement and Enjoyment in ATI Studies. *Educational Psychologist, 17*(2), 92-101. doi:10.1080/00461528209529247

Clark, R. E. (1989). When teaching kills learning: Research on mathemathantics. In H. N. Mandl, N. Bennett, E. de Corte, & H. F. Freidrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 2, pp. 1-22). London: Pergamon.

Clarke, C., Gage, N., Marx, R., Peterson, P., Stayrook, N., & Winne, P. (1979). A factorial experiment on teacher structuring, soliciting, and reacting. *Journal of Educational Psychology, 71*(4), 534-552.

Clarke, D., Keitel, C., & Shimizu, Y. (Eds.). (2006). *Mathematics classrooms in twelve countries: the insider's perspective*. Rotterdam, The Netherland: Sense Publishers.

Clarke, D., Mesiti, C., O'Keefe, C., Xu, L. H., Jablonka, E., Mok, I. A. C., & Shimizu, Y. (2007). Addressing the challenge of legitimate international comparisons of classroom practice. *International Journal of Educational Research, 46*(5), 280-293. doi:http://dx.doi.org/10.1016/j.ijer.2007.10.009

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitatvie data: complementary research strategies*. Thousand Oaks, CA: Sage.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. (6th ed.). London: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. (7th ed.). London: Routledge.

Coleman, J. S., Campbell, E., Hobson, C., MacPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington: DC: US Government Printing Office.

Cooper, D. C., & Schindler, P. S. (2001). *Business research methods*. (7th ed.). New York: McGraw-Hill.

Cooper, H. (1989). Synthesis of research on homework. *Educational leadership, 47*(3), 85-91.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003. *Review of Educational Research, 76*(1), 1-62. doi:10.3102/00346543076001001

Crawford, J., Gage, N. L., Corno, L., Stayrook, N., Mitman, A., Schunk, D., . . . Newman, R. (1978). *An experiment on teacher effectiveness and parent assisted instruction in the third grade*. Stanford, CA: Center for Educational Research at Stanford.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: the importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347-366. doi:10.1080/09243450600697242

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness : a contribution to policy, practice and theory in contemporary schools*. London: Routledge.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London: Routledge.

Croll, P. (1996). Teacher-pupil interaction in the classroom. In P. Croll & N. Hastings (Eds.), *Effective Primary Teaching* (pp. 14-28). London: David Fulton.

Croll, P., & Moses, D. (1985). *One in Five: the assessment and incidence of special edcuational needs*. London: Routledge and Kegan Paul.

Cuttance, P. (1998). Quality assurance reviews as catalyst for school improvement in Australia. In A. Hargreaves, A. Liberman, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change* (pp. 1135-1162). Dordrecht: Kluwer Publishers.

De Fraine, B., Van Damme, C., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: a multilevel latent growth approach. *Contemporary Educational Psychology, 32*(1), 132-150.

den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3-4), 407-442.

Department for Education and Employment (DfEE). (1997). *Excellence in Schools*. London: Her Majesty's Stationery Office.

Dewey, J. (1916). *Democracy and education*. New York: Macmillan.

Dexter, L. (1970). *Elite and specialized interviewing*. Evanston, IL: Northwestern University Press.

DfE. (2013). Schools in Southampton LA. Retrieved from: http://www.education.gov.uk/inyourarea/schools/lea_852_schools.shtml

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: an empirical test. *School Effectiveness and School Improvement, 11*(1), 57-79. Retrieved from http://www.tandfonline.com/doi/abs/10.1076/0924-3453%28200003%2911%3A1%3B1-A%3BFT057

Dunkin, M., & Biddle, B. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational leadership, 37*(1), 15-27.

Einsiedler, W., & Treinies, G. (1997). Effects of teaching methods, class effects, and patterns of cognitive teacher‐pupil interactions in an experimental study in primary school classes. *School Effectiveness and School Improvement, 8*(3), 327-353.

Emmer, E., Evertson, C., & Brophy, J. E. (1979). Stability of teacher effects in junior high classrooms. *American educational research journal, 16*(1), 71-75.

Emmer, E. T., Evertson, C., Sanford, J., & Clements, B. S. (1982). *Improving classroom management: an experimental study in junior high classrooms*.

Austin, TX: Research and Development Centre for Teacher Education, University of Texas.

Evertson, C., Anderson, C., Anderson, L., & Brophy, J. E. (1980). Relationships between classroom behaviours and student outcomes in junior high mathematics and English classes. *American educational research journal, 17*(1), 43-60.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviours, academic learning time, and student achievement: an overview. In C. Denham & A. Liberman (Eds.), *Time to Learn* (pp. 7-32). Washington, DC: National Institute of Education.

Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J., & Berliner, D. (1978). *Teaching behaviours, academic learning time and student achievement: final report of Phase III-B, Beginning Teacher Evaluation Study*. San Francisco: Far West Laboratory.

Flanders, N. (1970). *Analysing teacher behaviour*. Reading, MA: Addison-Wesley.

Flanders, N., & Simon, A. (1969). Teacher effectiveness. In R. L. Ebel (Ed.), *Encyclopedia of educational research* (4th ed., pp. 1423-1431). New York: Macmillan.

Flick, U. (1998). *An introduction to qualitative research*. London: Sage.

Fortune, J. (1967). *A study of the generality of presenting behaviours in teaching preschool children*. Memphis, TN: Memphis State University.

Fullan, M. (2007). *The new meaning of educational change*. (4th ed.). London: Routledge.

Gage, N. (1965). Desirable behaviors of teachers. *Urban Education, 1*(2), 85-95.

Gagné, E. D. (1985). *The cognitive psychology of school learning*. Boston: Little Brown and Co.

Gagné, R. (1970). *The conditions of learning*. New York: Hole, Rinehart and Winston.

Galton, M., Hargreaves, L., Comber, C., Wall, D., & Pell, A. (1999). *Inside the primary classroom: 20 years on*. London: Routledge.

Galton, M., & Simon, B. (Eds.). (1980). *Progress and performance in the primary classroom*. London: Reutledge & Kegan Paul.

Galton, M., Simon, B., & Croll, P. (1980). *Inside the Primary Classroom*. London: Routledge.

Gillham, B. (2008). *Observation techniques: structured to unstructured*. London: Continuum.

Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H. H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 Video Study. *Comparative Educational Review, 49*(3), 311-343.

Goldstein, H. (2003). *Multilevel models in educational and social research*. (3rd ed.). London: Edward Arnold.

Good, T. L., & Grouws, D. A. (1975). *Process-product relationship in fourth-grade mathematics classrooms*. Columbia: University of Missouri, College of Education.

Good, T. L., & Grouws, D. A. (1977). Teaching effects: a process-product study in fourth grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49-54.

Good, T. L., & Grouws, D. A. (1979). The Missouri Mathematics Effectiveness project: An experimental study in fourth-grade classrooms. *Journal of Educational Psychology, 71*(3), 355-362.

Good, T. L., Grouws, D. A., & Beckerman, T. (1978). Curriculum pacing: some empirical data in mathematics. *Journal of Curriculum Studies, 10*, 75-81.

Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.

Griffin, G. A., & Barnes, S. (1986). Using research findings to change school and classroom practice: results of an experimental study. *American educational research journal, 23(4)*, 572-586.

Grisay, A. (1996). *Evolution des acquis cognigifs et socio-affectifs des eleves au cours des annees de college (cognitive and socio-affective development during the years in the middle school)*. Liege: Universite de Liege.

Gu, L. (1994). *Theory of teaching experiment: The methodology and teaching principle of Qingpu*. [In Chinese 青浦实验的方法与教学原理研究]. Beijing, China: Educational Science Press.

Hallam, S. (2004). *Homework: The evidence*. London: Institute of Education, University of London.

Hampshire County Council. (2013). 2011 Census headline facts and figures Southampton. Retrieved from:

http://www3.hants.gov.uk/2011_census_southampton_summary_factsheet.pdf

Harris, A., Chapman, C., Muijs, D., Reynolds, D., Campbell, C., Creemers, B., . . . Weinstein, J. (2013). Getting lost in translation? An analysis of the international engagement of practitioners and policy-makers with the educational effectiveness research base. *School Leadership & Management, 33*(1), 3-19. doi:10.1080/13632434.2012.723622

Hart, K., Brown, M. L., Küchemann, D. E., Kerslake, D., Ruddock, G., & McCartney, M. (1981). *Children's understanding of mathematics: 11-16*. London: John Murray.

Hartas, D. (2010). Educational research and inquiry: key issues and debates. In D. Hartas (Ed.), *Educational research and inquiry: qualitative and quantitative approaches* (pp. 13-32). London: Continuum International Publishing Group.

Hattie, J. A. (2003). *Teachers make a difference: what is the research evidence?* Background paper to invited address presented at the 2003 ACER Research Conference, Carlton Crest Hotel, Melbourne, Australia, October 19-21, 2003.

Hennink, M. M. (2007). *International focus group research: a handbook for the health and social sciences*. Cambridge: Cambridge University Press.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., . . . Stigler, J. (2003). *Teaching mathematics in seven countries: results from the TIMSS 1999 Video Study*. Washington: National Center for Educational Statistics.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classrooms teaching: lessons from the TIMSS Video Study. *The Elementary School Journal, 101*(1), 3-20.

Hiebert, J., & Stigler, J. W. (2004). A world of difference. *Journal of Staff Development, 25*(4), 10-15.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal, 42*(2), 371-406.

Hill, P., & Rowe, K. (1995). Use of multi-level modelling in procedures for maximizing between school comparability of final year school-based assessments. *Multilevel Modelling Newsletter, 7*(1), 12-16.

Hill, P., & Rowe, K. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement, 7*(1), 1-34.

Hodgen, J., & Askew, M. (2007). Emotion, identity and teacher learning: becoming a primary mathematics teacher. *Oxford Review of Education, 33*(4), 469-487. doi:10.1080/03054980701451090

Hodgen, J., Küchemann, D. E., & Brown, M. (2014). Learning experiences designed to develop algebraic thinking: lessons from the ICCAMS project in England. In P. C. Toh, T. L. Toh, & B. Kaur (Eds.), *Learning Experiences that Promote Mathematics Learning* (pp. 171-185). Singapore: World Scientific.

Hodgkin, L. (2005). *A history of mathematics: From Mesopotamia to modernity*. Oxford: Oxford University Press.

Howe, C., Luthman, S., Ruthven, K., Mercer, N., Hofmann, R., Ilie, S., & Guardia, P. (2015). Rational number and proportional reasoning in early secondary school: towards principled improvement in mathematics. *Research in Mathematics Education, 17*(1), 38-56. doi:10.1080/14794802.2015.1019914

Huang, R. (2002). *Mathematics teaching in Hong Kong and Shanghai: A Classroom analysis from the perspective of variation*. Unpublished PhD thesis, University of Hong Kong.

Huang, R., Mok, I. A. C., & Leung, F. K. S. (2006). Practising in the Mathematics classrooms in China. In D. Clarke, C. Keitel, & Y. Shimizu (Eds.), *Mathematics Cassrooms in Twelve Countries: The Insider's Perspective*. Rotterdam, The Netherland: Sense Publishers.

Huggins, R., & Thompson, P. (2013). UK Competitiveness Index 2013. Retrieved from: http://www.cforic.org/downloads.php

Hunter, M., & Russell, D. (1981). Planning for effective instruction: lesson design *Increasing your teaching effectiveness*. Palo Alto, CA: The Learning Institute.

Husén, T. (1967). *International study of achievement in mathematics: a comparison of twelve countries*. (Vol. 1-2). Stockholm: Almqvist & Wiksell.

IEA. (2007). Mathematics concepts and mathematics items. Retrieved from: http://nces.ed.gov/timss/pdf/TIMSS4_Math_ConceptsItems.pdf

IEA. (2012). Brief history of IEA: 50 years of educational research. Retrieved from: http://www.iea.nl/brief_history.html

Ilonca, H. (2001). *The relationship between the use of representations and instructional discourse in mathematics tasks*. Paper presented at the the Anuual Meeting of the American Educational Research Association (AERA), Seattle, US.

Increasing Competence and Confidence in Algebra and Multiplicative Structures (ICCAMS) Maths. (2015). Survey results. Retrieved from: http://iccams-maths.org/survey/

Ingvarson, L., Beavis, A., Bishop, A., Peck, R., & Elsworth, G. (2004). *Investigation of effective mathematics teaching and learning in Australian secondary schools*. Canberra, Australian: Council for Educational Research.

Ivankova, N. V., & Kawamura, Y. (2010). Emerging trends in the utilisation of integrated designs in the social, behavioral, and health sciences. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral research* (2nd ed., pp. 581-612). London: SAGE.

Jacobs, J., Hiebert, J., Givvin, K. B., Hollingsworth, H., Garnier, H., & Wearne, D. (2006). Does eighth-grade mathematics teaching in the United States align with the NCTM "standards?" Results from the TIMSS 1996 and 1999 Video Studies. *Journal for Research in Mathematics Education, 37*(1), 5-32.

Jacobs, J. K., Hollingsworth, H., & Givvin, K. B. (2007). Video-based research made “easy”: methodological lessons learned from the TIMSS video studies. *Field Methods, 19*, 285-299.

James, H., Stigler, J. W., Jacobs, J., Givvin, K. B., Garnier, H., Smith, M., . . . Ronald, G. (2005). Mathematics teaching in the United States today (and tomorrow): results from the TIMSS 1999 Video Study. *Educational Evaluation and Policy Analysis, 27*(2), 111-132.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whole time has come. *Educational Research, 33*(7), 14-26.

Jones, K., Mooney, C., & Harries, T. (2002). Trainee primary teachers' knowledge of geometry for teaching. *Proceedings of the British Society for Research into Learning Mathematics, 22*(2), 95-100.

Jones, S. (1985). Depth interviewing. In R. Walker (Ed.), *Applied qualitative research* (pp. 45-55). Aldershot: Gower.

Keet, A. (2012). England. Retrieved from: http://www.england.org.za/

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Konstantopoulos, S., & Sun, M. (2013). Are teacher effects larger in small classes? *School Effectiveness and School Improvement, 25*(3), 312-328. doi:10.1080/09243453.2013.808233

Konu, A. I., Lintonen, T. P., & Autio, V. J. (2002). Evaluation of well-being in schools: a multilevel analysis of general subjective well-being. *School Effectiveness and School Improvement, 13*(2), 187-200.

Kounin, J. (1970). *Discipline and group management in classrooms*. New York: Holt, Rineheart & Winston.

Krantz, S. G. (2010). *An episodic history of mathematics: mathematical culture through problem solving*. USA: The Mathematical Association of America.

Krueger, R. A. (1998). *Developing questions for focus groups*. Thousand Oaks: Sage Publications.

Krueger, R. A., & Casey, M. A. (2009). *Focus groups : a practical guide for applied research*. (4th ed.). Thousand Oaks, CA: SAGE.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic Findings on Grouping Programs. *Gifted Child Quarterly, 36*(2), 73-77. doi:10.1177/001698629203600204

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*, 12-23.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Lampert, M. (1988). What can research on teacher education tell us about improving the quality of mathematics education? *Teaching and Teacher Education: An International Journal of Research and Studies, 3(4)*, 157-170.

Land, M. (1979). Low-inference variables and teacher clarity: effects on student concept learning. *Journal of Educational Psychology, 71*, 795-799.

Lapointe, A. E., Mead, N. A., & Askew, J. M. (1992). *Learning mathematics: an international assessment of educational progress*. Princeton: ETS.

Lapointe, A. E., Mead, N. A., & Phillips, G. W. (1989). *A world of differences: an international assessment of mathematics and science*. Princeton: ETS.

LeCompte, M., & Preissle, J. (1993). *Ethnography and qualitative design in educational research*. (2nd ed.). London: Academic Press.

LessonLab. (2012). Data Collection Manual. Retrieved from: http://timssvideo.com/sites/default/files/Data Collection Manual.pdf

Leung, F. K. S. (2005). Some characteristics of East Asian mathematics classrooms based on data from the TIMSS 1999 Video Study. *Educational Studies in Mathematics, 60*(2), 199-215.

Lichtman, M. (2010). *Qualitative research in education: a user's guide*. (2nd ed.). London: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, Calif.: Sage Publications.

Luyten, H. (1998). School effectiveness and student achievement, consistent across subjects? Evidence from Dutch elementary and secondary education. *Educational research and Evaluation, 4*(4), 281-306.

Luyten, H. (2003). The size of school effects compared to teacher effects: an overview of the research literature. *School Effectiveness and School Improvement, 14*(1), 31-51. doi:10.1076/sesi.14.1.31.13865

Luyten, H., & de Jong, R. (1998). Parallel classes: differences and similarities. Teacher effects and school effects in secondary schools. *School Effectiveness and School Improvement, 9*(4), 437-473. doi:10.1080/0924345980090405

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

MacNamara, D. R., & Waugh, D. G. (1993). Classroom organisation: a discussion of grouping strategies in the light of the 'Three Wise Men's' report. *School Organisation, 13*(1), 41-50.

Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (Eds.). (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

McCaleb, J., & White, J. (1980). Critical dimensions in evaluating teacher clarity. *Journal of Classroom Interaction, 15*(2), 27-30.

Meijnen, G. W., Lagerweij, N. W., & Jong, P. F. (2003). Instruction characteristics and cognitive achievement of young children in elementary schools. *School Effectiveness and School Improvement, 14*(2), 159-187. doi:10.1076/sesi.14.2.159.14224

Merton, R. K., & Kendall, P. L. (1946). The focused interview. *American Journal of Sociology, 51*, 541-557.

Miller, G. A. (1956). The magic number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Minichiello, V., Aroni, R., Timewell, E., & Alexander, L. (1990). *In-depth interviewing: researching people*. Melbourne: Longman Cheshire.

Mok, I. A. C. (2009). In search of an exemplary mathematics lesson in Hong Kong: an algebra lesson on factorization of polynomials. *ZDM - The International Journal on Mathematics Education, 41*(3), 319-332.

Morgan, D. L. (1988). *Focus groups as qualitative research*. Beverly Hills, CA: SAGE.

Mortimore, P. (1991). School effectiveness research: which way at the crossroads? *School Effectiveness and School Improvement, 2*(3), 213-229.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School Matters*. Wells: Open books.

Muijs, D. (2011). *Doing quantitative research in education with SPSS*. (2nd ed.). London: SAGE.

Muijs, D., & Dunne, M. (2010). Setting by ability–or is it? A quantitative study of determinants of set placement in English secondary schools. *Educational Research, 52*(4), 391-407.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231-256. doi:10.1080/09243453.2014.885451

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the Mathematics Enhancement Programme (Primary). *School Effectiveness and School Improvement, 11*(3), 273-303. Retrieved from http://www.tandfonline.com/doi/abs/10.1076/0924-3453%28200009%2911%3A3%3B1-G%3BFT273

Muijs, D., & Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics: a longitudinal study. *Educational Research and Evaluation: An International Journal on Theory and Practice, 9*(3), 289-314.

Muijs, D., & Reynolds, D. (2011). *Effective teaching: Evidence and practice*. (3rd ed.). London: SAGE.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Science Achievement in the Primary School Years. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation and Educational Policy, Boston College.

Mullis, I. V. S., Martin, M. O., & Foy, P. (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.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 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.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999: international mathematics reports: finding from IEA's Repeat of Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: The International Study Center, Boston College.

Munro, J. (1999). Learning more about learning improves teacher effectiveness. *School Effectiveness and School Improvement, 10*(2), 151-171.

Nanjing Education Bureau. (2011). *Education development in Nanjing in 2011*. Retrieved from http://www.njedu.gov.cn/default.php?mod=article&do=detail&tid=669644

National Bureau of Statistics of China. (2011). *2010 Census of China*. Retrieved from http://www.stats.gov.cn/tjgb/rkpcgb/qgrkpcgb/t20110428_402722232.htm

National Bureau of Statistics of China. (2013). *Annual data of major cities in China*. Retrieved from: http://data.stats.gov.cn/workspace/index?m=csnd

Nuffield Foundation. (1994). *History of Mathematics*. Singapore: Longman.

OECD. (2001). *Knowledge and skills for life: first results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD Publishing.

OECD. (2009a). *Creating effective teaching and learning environments: first results from TALIS*. Paris: OECD Publishing.

OECD. (2009b). Teaching and learning international survey (TALIS): teacher questionnaire. Retrieved from: http://www.oecd.org/edu/school/43081350.pdf

OECD. (2010). *PISA 2009 results: what students know and can do – student performance in reading, mathematics and science*. (Vol. I). Paris: OECD Publishing.

OECD. (2013). *PISA 2012 results: what students know and can do - student performance in mathematics, reading and science*. (Vol. I). Paris: OECD Publishing.

OECD. (2014). *TALIS 2013 results: an International perspecitve on teaching and learning*. doi:http://dx.doi.org/10.1787/9789264196261-en

Office for National Statistics. (2013a). 2011 UK censuses. Retrieved from: http://www.ons.gov.uk/ons/guide-method/census/2011/uk-census/index.html

Office for National Statistics. (2013b). The countries of the UK. Retrieved from: http://web.archive.org/web/20081220225201/http://www.statistics.gov.uk/geography/uk_countries.asp

Oldroyd, D. (1986). *The arch of knowledge: an introductory study of the history of the philosophy and methodology of science*. New York: Methuen.

Opdenakker, M.-C., Maulana, R., & den Brok, P. (2012). Teacher–student interpersonal relationships and academic motivation within one school year: developmental changes and linkage. *School Effectiveness and School Improvement, 23*(1), 95-119.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Pinter.

Oxford Dictionaries. (n.d.). The definition of pedagogy in English. Retrieved from: http://www.oxforddictionaries.com/definition/english/pedagogy

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: the importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111-140.

Park, K., & Leung, F. K. S. (2006). Mathematics lessons in Korea: Teaching with systematic variation. In D. Clarke, C. Keitel, & Y. Shimizu (Eds.), *Mathematics Classrooms in Twelve Countries: The Insider's Perspective*. Rotterdam, The Netherland: Sense Publishers.

Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills, CA: Sage.

Peng, W. J., McNess, E., Thomas, S., Wu, X. R., Zhang, C., Li, J. Z., & Tian, H. S. (2014). Emerging perceptions of teacher quality and teacher development in China. *International Journal of Educational Development, 34*, 77-89. doi:http://dx.doi.org/10.1016/j.ijedudev.2013.04.005

Piaget, J. (1950). *The psychology of intelligence*. London: Routledge & Kegan Paul.

Piaget, J. (1962). *Play, dreams and imitation in childhood*. London: Routledge and Kegan Paul.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. London: Routledge & Kegan Paul.

Pollard, A., Broadfoot, P., Croll, P., Osborn, M., & Abbott, D. (1994). *Changing English primary schools? The impact of the Education Reform Act at Key Stage 1*. London: Cassell.

Potari, D., Sakonidis, H., Chatzigoula, R., & Manaridis, A. (2010). Teachers' and researchers' collaboration in analysing mathematics teaching: A context for professional reflection and development. *Journal of Mathematics Teacher Education, 13*(6), 473-485. doi:10.1007/s10857-010-9161-9

PRC Gov. (2013). China profile. Retrieved from: http://www.gov.cn/test/2005-08/11/content_27116.htm

Punch, K. F. (2005). *Introduction to social research: quantitative and qualitative approaches*. (2nd ed.). London: Sage.

Punch, K. F. (2009). *Introduction to research methods in education*. London: Sage.

Reid, E. R. (1978-1982). *The Reader Newsletter*. Salt Lake City, Utah: Examplary Center for Reading Instruction.

Reynolds, D. (1976). The delinquent school. In M. Hammersley & P. Woods (Eds.), *The process of schooling* (pp. 217-219). London: Routledge & Kegan Paul.

Reynolds, D. (1995). The effective school: an inaugural lecture. *Evaluation and Research in Education, 9*(2), 57-73.

Reynolds, D. (2010). *Failure free education? The past, present and future of school effectiveness and school improvement*. London: Routledge.

Reynolds, D. (Ed.) (1985). *Studying school effectiveness*. London: Falmer.

Reynolds, D., Bollen, R., Creemers, B. P. M., Hopkins, D., Stoll, L., & Lagerweij, N. (1996). *Making good schools: linking school effectiveness and school improvement*. London: Routledge.

Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., & Schaffer, G. (Eds.). (2002a). *World class schools: international perspectives on school effectiveness*. London: Routledge.

Reynolds, D., & Farrell, S. (1996). *Worlds apart? A review of international surveys of educational achievement involving England*. London: H.M.S.O.

Reynolds, D., Sammons, P., de Fraine, B., Townsend, T., & van Damme, J. (2011). *Educational effectiveness research: a state of the art review*. Paper presented at the The Annual Meeting of the International Congress for School Effectiveness and Improvement, Cyprus.

Reynolds, D., Stringfield, S., Teddlie, C., & Creemers, B. P. M. (2002b). The intellectual and policy context. In D. Reynolds, B. P. M. Creemers, S. Stringfield, C. Teddlie, & G. Schaffer (Eds.), *World class schools:*

international perspective on school effectiveness (pp. 3-14). London: Routledge Falmer.

Reynolds, D., & Teddlie, C. (2000). The processes of school effectiveness. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 134-159). London: Falmer press.

Rice, S. A. (Ed.) (1931). *Methods in social science*. Chicago: University of Chicago Press.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice: a guide for social science students and researchers*. London: Sage.

Robitaille, D. F., & Garden, R. A. (Eds.). (1989). *The IEA Study of Mathematics II: contexts and outcomes of school mathematics*. Oxford: Pergamon Press.

Rosenshine, B. (1971). *Teaching behaviours and student achievement*. London: National Foundation and Educational Research.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 122-183). Chicago: Rand McNally.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376-391). New York: Macmillan.

Rossella, S., & Alessandra, B. (2005). Mathematics teaching in Italy: a cross-cultural video analysis. *Mathematical Thinking & Learning, 7*(4), 291-312.

Rowland, T. (2010). Foundation knowledge for teaching: Contrasting elementary and secondary mathematics. In V. Durand-Guerrier, S. Soury-Lavergne, &

F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education* (pp. 1841-1850). Lyon: CERME.

Ruddock, G., Sainsbury, M., Clausen-May, T., Vppula, H., Mason, K., & Patterson, E. W. (2008). *Comparison of the core primary curriculum in England to those of other high performing countries*. London: DCSF.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. w. S., A.,. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Boston, MA: Harvard University Press.

Sammons, P. (1999). *School effectiveness: coming of age in the 21st century*. Liss: Swets & Zeitlinger.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: effective departments and effective schools*. London: Paul Chapman.

Scheerens, J., & Bosker, R. (1997). *The foundations of school effectiveness*. Oxford: Pergamon Press.

Scheerens, J., Vermeulen, C. J. A. J., & Pelgrum, W. J. (1989). Generalizability of instrucitonal and school effectiveness indicators across nations. *International Journal of Educational Research, 13*(7), 789-799.

Schmidt, W. H., Jorde, D., Cogan, L. S., Barrier, E., Gonzalo, I., Moser, U., . . . Wolfe, R. G. (1996). *Characterizing pedagogical flow: an investigation of mathematics and science teaching in six countries*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogen, L. S., & Wolfe, R. G. (2001). *Why schools matter: a cross-*

national comparison of curriculum and learning. San Fransisco: Jossey-Bass.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner toward a new design for teaching and learning in teh professions*. San Francisco, CA: Jossey-Bass.

Sekiguchi, Y. (2006). Coherence of mathematics lessons in Japanese eighth-grade classrooms. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlíková (Eds.), *Proceedings of 30 th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 81-88).

Shimizu, Y. (2007). Explicit linking in the sequence of consecutive lessons in mathematics classroom in Japan *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 177-184). Seoul, South Korea: PME.

Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher, 15(2)*, 4-14.

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard educational review, 57(1)*, 1–22.

Smith, D. J., & Tomlinson, S. (1989). *The school effect: a study of multi-racial comprehensives*. London: Policy Studies Institute.

Smith, L., & Cotton, M. (1980). Effect of lesson vagueness and discontinuity on student achievement and attitudes. *Journal of Educational Psychology, 72*, 670-675.

Smith, L., & Land, M. (1981). Low-inference verbal behaviours related to teacher clarity. *Journal of Classroom Interaction, 17*(1), 37-42.

Soar, R. S. (1966). *An integrative approach to classroom learning*. Philadelphia: Temple University.

Soar, R. S. (1973). *Follow Through classroom process measurement and pupil growth*. Gainesville: University of Florida.

Soar, R. S. (1977). An integration of findings from four studies of teacher effectiveness. In G. Borich & K.S. Fenton (Eds.), *The appraisal of teaching: concepts and process* (pp. 146-161). New York: Addison-Wesley.

Soar, R. S., & Soar, R. M. (1972). An empirical analysis of selected Follow Through programmes: an example of a process approach to evaluation. In I. Gordon (Ed.), *Early childhood education* (pp. 229-259). Chicago, IL: National Society for the Study of Education.

Soar, R. S., & Soar, R. M. (1973). *Classroom behaviour, pupil characteristics and pupil growth for the school year and the summer*. Gainesville: University of Florida, Institute for Development of Human Resources.

Soar, R. S., & Soar, R. M. (1978). *Setting variables, classroom interaction, and multiple pupil outcomes*. Washington, D.C.: National Institute of Education.

Southampton City Council. (2013). Southampton city statistics and research. Retrieved from: http://www.southampton.gov.uk/living/statsresearch/

Stallings, J., Cory, R., Fairweather, J., & Needles, M. (1977). *Early childhood education classroom evaluation*. Menlo Park, CA: SRI International.

Stallings, J., Cory, R., Fairweather, J., & Needles, M. (1978). *A study of basic reading skills taught in secondary schools*. Menlo Park, CA: SRI International.

Stallings, J., & Kaskowitz, D. (1974). *Follow Through classroom observation evaluation 1972-1973*. Stanford: Stanford Research Institute.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: why our schools are failing, and what we can learn from Japanese and Chinese education*. New York: Summit Books.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS Videotape Classroom Study: methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. Washington, DC: National Centre for Education Statistics.

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

Stigler, J. W., & Perry, M. (1988). Mathematics learning in Japanese, Chinese, and American classrooms. *New Directions for Child and Adolescent Development, 1988*(41), 27-54.

Stockero, S. L. (2008). Using a video-based curriculum to develop a reflective stance in prospective mathematics teachers. *Journal of Mathematics Teacher Education, 11*(5), 373-394. doi:10.1007/s10857-008-9079-7

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE.

Tashakkori, A., & Teddlie, C. (2010). *SAGE handbook of mixed methods in social & behavioral research*. (2nd ed.). Los Angeles: SAGE.

Teddlie, C., Creemers, B. P. M., Kyriakides, L., Muijs, D., & Fen, Y. (2006). The international system for teacher observation and feedback: evolution of an international study of teacher effecitveness constructs. *Educatonal Research ad Evaluation, 12*(6), 561-582.

Teddlie, C., & Liu, S. (2008). Examining teacher effectiveness within differentially effective primary schools in the People's Republic of China. *School Effectiveness and School Improvement, 19*(4), 387-407. doi:10.1080/09243450802535182

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

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: lessons learned from a ten year study of school effects*. New York: Teachers College Press.

Teodorović, J. (2011). Classroom and school factors related to student achievement: what works for students? *School Effectiveness and School Improvement, 22*(2), 215-236. doi:10.1080/09243453.2011.575650

The World Bank. (2013). GDP per capita. Retrieved from: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD/countries?order=

wbapi_data_value_2012 wbapi_data_value wbapi_data_value-last&sort=desc&display=default

Tikunoff, W., Berliner, D., & Rist, R. (1975). *An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample*. San Francisco: Far West Laboratory.

Tizard, B., Blatchford, P., Burke, J., Farquhar, C., & Plewis, I. (1988). *Young children at school in the inner city*. Hove: Lawraence Erlbaum Associate Ltd.

Tobin, J. J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in three cultures revisited: China, Japan, and the United States*. Chicago: The University of Chicago Press.

Tobin, J. J., Wu, D. Y. H., & Davidson, D. H. (1989). *Preschools in three cultures: Japan, China, and the United States*. New Haven: Yale University Press.

Ulewicz, M., & Beatty, A. (Eds.). (2001). *Power of video technology in international comparative research in education*. Washington: National Academies Press.

Usher, R. (1996). A critique of the neglected epistemological assumptions of educational research. In D. Scott & R. Usher (Eds.), *Understanding educational research* (pp. 9-32). London: Routledge.

Van Damme, J., Opdenakker, M., Van Landeghem, G., De Fraine, B., Pustjens, H., & Van de gaer, E. (2006). *Educational effectiveness: an introduction to international and Flemish research on school, teachers and classes*. Leuven: Acco.

Van de gaer, E., De Fraine, B., Van Damme, J., De Munter, A., & Onghena, P. (2009). School effects on the development of motivation toward learning tasks and the development of academic self-concept in secondary school. *School Effectiveness and School Improvement, 20*(2), 235-253.

van de Grift, W. J. C. M. (2013). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement, 25*(3), 295-311. doi:10.1080/09243453.2013.794845

Van der Werf, G., Opdenakker, M.-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel latent growth curve approach. *School Effectiveness and School Improvement, 19*(4), 447-462.

Van Landeghem, G., Van Damme, J., Opdenakker, M.-C., De Fraine, B., & Onghena, P. (2002). The effect of schools and classes on noncognitive outcomes. *School Effectiveness and School Improvement, 13*(4), 429-451.

Verachtert, P., Van Damme, J., Onghena, P., & Ghesquiere, P. (2009). A seasonal perspective on school effectiveness: evidence from a Flemish longitudinal study in kindergarten and first grade. *School Effectiveness and School Improvement, 20*(2), 215-233.

Voight, J. (1994). Negotiation of mathematical meaning and learning mathematics. *Educational Studies in Mathematics, 26*(2-3), 275-298.

Walker, D., & Schaffarzick, J. (1974). Comparing curricula. *Review of Educational Research, 44*, 83-111.

Walshaw, M. (2010). Mathematics pedagogical change: rethinking identity and reflective practice. *Journal of Mathematics Teacher Education, 13*(6), 487-497. doi:10.1007/s10857-010-9163-7

Weber, G. (1971). *Inner city children can be taught to read: Four successful schools*. Washington DC: Council for Bacis Education.

Weber, M. (1949). *The methodology of the social sciences*. Illinois, the US: The Free Press of Glengoe.

Wikipedia. (2013). List of countries by GDP (nominal) per capita. Retrieved from: http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita

Wilkinson, J. (2000). Direct observation. In G. M. Breakwell, S. Hammond, & C. Fife-Shaw (Eds.), *Research methods in psychology* (pp. 224-238). London Sage.

Willms, J. D. (2000). Monitoring school performance for standards-based reform. *Evaluation and Research in Education, 14*(3-4), 237-253.

Wragg, E. C. (1999). *An introduction to classroom observation*. (2nd ed.). London: Routledge.

APPENDICES

List of Appendices

APPENDIX A: Ethics Approval, Info Sheets & Consent Form

From: "ERGO"
Date: Monday, 5 November 2012 15:40:05 Greenwich Mean Time
To: Z.Miao@soton.ac.uk
Cc:
Subject: **Research Governance Feedback on your Ethics Submission (Ethics ID:4266)**

Submission Number 4266:
Submission Title The effectiveness of mathematics teaching: A cross-national investigation in primary schools in China and the UK:
The Research Governance Office has reviewed and approved your submission

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment) or external ethics review (e.g. NRES).The following comments have been made:

"Many thanks for making the changes, I can confirm this study will be covered by the University insurance programme"

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL date-sent 1352130005 flags 33815937 original-mailbox ews://zm3g10@www.outlook.soton.ac.uk/Inbox remote-id AAMkAGM1ZmM4Yjc5LWI5YmYtNDM2ZC1hMjUwLWUzYWFkNWJjYzViYQBGAAAAAADbRLoOLxasRLyFLVD3PB+TBwA1HuqzNlr/S65W2NsZJsevAMKX6Gr+AADYtyo4R+kZQZiriT4K7c+IAAAACwT7AAA= sender ERGO <ergo@soton.ac.uk> subject Research Governance Feedback on your Ethics Submission (Ethics ID:4266) to <Z.Miao@soton.ac.uk>

UNIVERSITY OF Southampton

Participant Information Sheet for Parents/Guardians

Study Title: The effectiveness of mathematics teaching: A cross-national investigation in primary schools in China and the UK

Researcher: Zhenzhen Miao **Ethics number**: 4266

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?
I am a postgraduate student from China and am currently studying in Southampton Education School, University of Southampton in the UK. This is a research project that I conduct towards a qualification of PhD degree.

The project is designed to address the issues in relation to the effectiveness of mathematics teaching at the primary phase across two countries and two cultures, i.e. China and the UK. In order to collect reliable data, research methods will include two standardised mathematics tests for pupils, classroom observations recorded by camcorders, interviews and focus groups with teachers who deliver lessons, with the teachers' colleagues in the same regions, and with the teachers' colleagues in the other involved country, one questionnaire to teachers about their backgrounds, and another questionnaire to students about their perceptions of mathematics teaching and learning.

Why have I been chosen?
Your child's teacher is teaching right at the year-level that fits the targeted level of the project and the local education authority has recommended your child's school which in turn has recommended your child's teacher as a candidate to have one of her/his mathematics lessons observed and studied. Once all consent forms have been obtained, your child will be in the class with other children while their teacher is delivering a mathematics lesson.

What will happen to me if I take part?

- The lesson will be observed and video-recorded, so your child and the rest of the class will be in the video that the camera captures from the beginning to the end of the lesson.

- Your child is also expected to fill in a questionnaire about children's perception of mathematics teaching and learning, and to complete two mathematics tests, one in February 2013 and the other in June 2013.

Are there any benefits in my taking part?
There might be no benefit to the individual, but your child's participation will promote the development of research on teaching effectiveness in mathematics across countries.

Are there any risks involved?
There're no risks about your child's participation in this project. During the study, only the student and the supervisor have access to the data for the purpose of research.

Digital data will be kept by the researcher in a password-protected computer during the study and in the university database after the study.

Paper-based questionnaires and test papers will all be stored in a coded suitcase during the data-collection process, be stored in a locked storage during the data-analysis process, and be safely destroyed after the study.

Will my participation be confidential?

- The data collected will be only used for this project and will be stored either in a password-protected computer or in a locked storage by the researcher during the study. Digital data will be stored in the university database and paper-based data will be safely destroyed when the project is completed. Any further use of the data will need your additional consent on it.

- All names of participants and participating schools will be anonymised in the thesis and any publication of the study, and your child's confidentiality will be put at the highest priority in this project.

What happens if I change my mind?

You have the right to decide whether or not your child is going to participate the project, and you can withdraw at any time without your and your child's legal rights being affected.

What happens if something goes wrong?

If you have any concerns or complaints about how this research is conducted, you may contact:

Dr Martina Prude, Head of Research and Governance

University of Southampton, UK

+44 2380 595058

Mad4@soton.ac.uk

Where can I get more information?

If you have any further questions once you have read this information sheet, please get in touch with me using the following details:

Zhenzhen Miao
Southampton Education School
The University of Southampton
Building 32, University Road
Southampton
SO17 1BJ
U.K.
Email: Z.Miao@soton.ac.uk

UNIVERSITY OF Southampton

CONSENT FORM FOR PARENTS/GUARDIANS

Study title: The effectiveness of mathematics teaching: A cross-national investigation in primary schools in China and the UK

Researcher name: Zhenzhen Miao

Ethics reference: 4266

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (dated 02/11/2012 version ZM1.2) and have had the opportunity to ask questions about the study.	☐
I agree for my child to take part in this research project and agree for my child's data to be used for the purpose of this study.	☐
I agree for my child to complete two standardised mathematics tests along with other children in the same class.	☐
I agree for my child to be videoed in their mathematics lessons and agree for the videos containing my child's images to be used for the purpose of this study.	☐
I agree for my child to complete a questionnaire on their perceptions of mathematics teaching and learning.	☐
I understand my child's participation is voluntary and I may withdraw at any time without my and my child's legal rights being affected.	☐

UNIVERSITY OF Southampton

Data Protection

I understand that information collected about my child during my child's participation in this study will be kept safely. Digital data will be kept by the researcher in a password-protected computer during the study and in the university database after the study. Paper-based questionnaires and test papers will all be stored in a coded suitcase during the data-collection process, be stored in a locked storage during the data-analysis process, and be safely destroyed after the study. During the study, only the researcher and her supervisor have access to the data. I understand that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of Parents/Guardians (print name)...

Signature of Parents/Guardians...

Name of the Child (print name)...

Signature of the Child...

Date...

UNIVERSITY OF Southampton

Participant Information Sheet for Teachers

Study Title: The effectiveness of mathematics teaching: A cross-national investigation in primary schools in China and the UK

Researcher: Zhenzhen Miao **Ethics number**: 4266

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

I am a postgraduate student from China and am currently studying in Southampton Education School, University of Southampton in the UK. This is a research project that I conduct towards a qualification of PhD degree.

The project is designed to address the issues in relation to the effectiveness of mathematics teaching at the primary phase across two countries and two cultures, i.e. China and the UK. In order to collect reliable data, research methods will include two standardised mathematics tests for pupils, classroom observations recorded by camcorders, interviews and focus groups with teachers who deliver lessons, with the teachers' colleagues in the same regions, and with the teachers' colleagues in the other involved country, one questionnaire to teachers about their backgrounds, and another questionnaire to students about their perceptions of mathematics teaching and learning.

Why have I been chosen?

You are chosen because you are a mathematics teacher who is teaching right at the year-level that fits the targeted level of the project and the local education authority in your region has recommended your school which in turn recommended you as a potential participant.

What will happen to me if I take part?

- One of your lessons will be observed and video-recorded.
- You'll be interviewed after the lesson while watching the video of your lesson.
- You're also expected to complete a questionnaire about your background.
- You will be invited to join a focus group to discuss a lesson with several other native teachers, while watching the video of a lesson delivered by a native teacher.
- You will also be invited to join a focus group to discuss a lesson with several other native teachers, while watching the video of a lesson delivered by a foreign teacher from China/England (deleted as applicable).

Are there any benefits in my taking part?

There might be no benefit to the individual, but your participation will increase the professional communications on the teaching of mathematics across countries and also promote the development of research on teaching effectiveness in mathematics across countries.

Are there any risks involved?

UNIVERSITY OF Southampton

There're not any risks about your participation in this project. During the study, only the student and the supervisor have access to the data for the purpose of research.

Digital data will be kept by the researcher in a password-protected computer during the study and in the university database after the study.

Paper-based questionnaires and test papers will all be stored in a coded suitcase during the data-collection process, be stored in a locked storage during the data-analysis process, and be safely destroyed after the study.

Will my participation be confidential?

- The data collected will be only used for this project and will be stored either in a password-protected computer or in a locked storage by the researcher during the study. Digital data will be stored in the university database and paper-based data will be safely destroyed when the project is completed. Any further use of the data will need your additional consent on it.

- All names of participants and participating schools will be anonymised in the thesis and any publication of the study, and your confidentiality will be put at the highest priority in this project.

What happens if I change my mind?

You are the only one who has the right to decide whether to participate or not the project. Therefore, you can withdraw any time you want without your legal rights being affected.

What happens if something goes wrong?

If you have any concerns or complaints about how this research is conducted, you may contact:

Dr Martina Prude, Head of Research and Governance

University of Southampton, UK

+44 2380 595058

Mad4@soton.ac.uk

Where can I get more information?

If you have any further questions once you have read this information sheet, please get in touch with me using the following details:

Zhenzhen Miao
Southampton Education School
The University of Southampton
Building 32, University Road
Southampton
SO17 1BJ
U.K.
Email: Z.Miao@soton.ac.uk

UNIVERSITY OF Southampton

CONSENT FORM FOR TEACHERS

Study title: The effectiveness of mathematics teaching: A cross-national investigation in primary schools in China and the UK

Researcher name: Zhenzhen Miao

Ethics reference: 4266

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (dated 16/10/2012 version ZM1.1) and have had the opportunity to ask questions about the study.	☐
I agree to take part in this research project and agree for my data to be used for the purpose of this study	☐
I agree for my lesson to be observed and video-recorded for analysis and agree for this lesson video to be observed and discussed by both Chinese and English colleagues.	☐
I agree to be interviewed to make comments on my own lesson video and agree for the interview to be audio-recorded for the research analysis.	☐
I agree to fill in a questionnaire for the purpose of the research.	☐
I agree to be interviewed in a focus group with teachers from my country, whose mathematics lessons have been observed, to watch and comment on two mathematics lessons, one taught by a native teacher and the other by a foreign teacher, and I agree for the discussion in the focus group to be audio-recorded for the purpose of this study.	☐

UNIVERSITY OF Southampton

I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected

Data Protection

I understand that information collected about me during my participation in this study will be kept safely. Digital data will be kept by the researcher in a password-protected computer during the study and in the university database after the study. Paper-based questionnaires and student test papers will all be stored in a coded suitcase during the data-collection process, be stored in a locked storage during the data-analysis process, and be safely destroyed after the study. During the study, only the researcher and her supervisor have access to the data. I understand that my information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

.

Name of participant (print name)……………………………………………………

Signature of participant…………………………………………………………..

Date…………………………………………………………………………………

APPENDIX B: Two Observation Systems – OTL & ISTOF

System A: OTL

Opportunity to Learn (OTL) designed by the ISERP Team (Reynolds et al., 2002)
(EMT adapted version)

Activity code	Time	Descriptive notes	Time on task (every 5 mins)	
			Category	Number
			Time: On task: Off task: Waiting: Out of class:	
			Time: On task: Off task: Waiting: Out of class:	
			Time: On task: Off task: Waiting: Out of class:	

Activity Key: 1 = Whole class interactive
2 = Whole class lecture
3 = individual/group work
4 = Classroom management
5 = Partial class interactive

a = Calculators
b = Collaborative

System B: ISTOF (EMT adapted version)

ISTOF Teacher Observation Protocol
(the EMT adapted version)

The ISTOF Teacher Observation Protocol was designed by an international team of experts in the area of teacher effectiveness to measure observable teacher behaviours consistent with effective classroom teaching. Each observation should be conducted in a regular classroom setting and last for an entire class period (typically 40 - 50 minutes). The observer should rate each item according to the following rating scale.

5 – Strongly Agree
4 – Agree Somewhat
3 - Neutral
2 – Disagree Somewhat
1 - Strongly Disagree
NA - Not applicable/unable to observe

ISTOF 1: ASSESSMENT AND EVALUATION

(1) Indicator #1.1: The teacher gives explicit, detailed and constructive feedback

Item #1: *The teacher makes explicitly clear why an answer is correct or not.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #2: *The teacher provides appropriate feedback to the answers given by the students.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

(2) Indicator #1.2: Assessment is aligned with goals and objectives

Item #3: *Assignments given by the teacher are clearly related to what students learned.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #4: *The teacher explains how assignments are aligned to the learning goals of the lesson.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF 2: CLARITY OF INSTRUCTION

(3) Indicator #2.1: The teacher shows good communication skills

Item #5: *The teacher regularly checks for understanding.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #6: *The teacher communicates in a clear and understandable manner.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

(4) Indicator #2.2: Clear explanation of purpose

Item #7: *The teacher clarifies the lesson objectives at the start of the lesson.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #8: *The teacher asks students to identify the reasons why specific activities take place in the lesson.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(5) Indicator #2.3: Lessons are well structured

Item #9: *The teacher presents the lesson with a logical flow that moves from simple to more complex concepts.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #10: *The teacher implements the lesson smoothly moving from one stage to another with well-managed transition points.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

ISTOF 3: INSTRUCTIONAL SKILLS

(6) Indicator #3.1: The teacher is able to engage students

Item #11: *The teacher provides sufficient wait time and response strategies to involve all types of learners.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #12: *The teacher gives assignments that stimulate all students to active involvement.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(7) Indicator #3.2: The teacher possesses good questioning skills

Item #13: *The teacher poses questions which encourage thinking and elicit feedback.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #14: *The length of the pause following questions varies according to the difficulty level of questions (e.g., a question calling for application of abstract principles requires a longer pause than a factual question).*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

(8) Indicator #3.3: The teacher uses various teaching methods and strategies

Item #15: *The teacher uses a variety of instructional strategies during the class period.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF 4: PROMOTING ACTIVE LEARNING AND DEVELOPING METACOGNITIVE SKILLS

(9) Indicator #4.1: The teacher helps pupils develop problem-solving and meta-cognitive strategies

Item #16: *The teacher invites students to use strategies which can help them solve different types of problems.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #17: *The teacher invites students to explain the different steps of the problem solving strategy which they are using.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

Item #18: *The teacher explicitly provides instruction in problem-solving strategies.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(10) Indicator #4.2: The teacher gives students opportunities to be active learners

Item #19: *The teacher encourages students to ask one another questions and to explain their understanding of topics to one other.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #20: *The teacher gives students the opportunity to correct their own work.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(11) Indicator #4.3: The teacher fosters critical thinking in students.

Item #21: *The teacher motivates the students to think about the advantages and disadvantages of certain approaches.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

Item #22: *The teacher asks the students to reflect on the solutions/answers they gave to problems or questions.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #23: *The teacher invites the students to give their personal opinion on certain issues.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(12) Indicator #4.4: The teacher connects material to students' real world experiences

Item #24: *The teacher systematically uses material and examples from the students' daily life to illustrate the course content.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #25: *Students are invited to give their own examples.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

ISTOF 5: CLASSROOM CLIMATE

(13) Indicator #5.1: All students are valued.

Item #26: *The teacher demonstrates genuine warmth and empathy toward all students in the classroom.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #27: *The teacher shows respect for the students in both in his/her behaviour and use of language.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(14) Indicator #5.2: The teacher initiates active interaction and participation.

Item #28: *The teacher creates purposeful activities that engage every student in productive work.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #29: *The teacher's instruction is interactive (lots of questions and answers).*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

(15) Indicator #5.3: The teacher interacts with all students

Item #30: *The teacher gives turns to and/or involves those students who do not voluntarily participate in classroom activities.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #31: *The teacher seeks to engage all students in classroom activities.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(16) Indicator #5.4: The teacher communicates high expectations

Item #32: *The teacher praises children for effort towards realizing their potential.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #33: *The teacher makes clear that all students know that he/she expects their best efforts in the classroom.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

ISTOF 6: CLASSROOM MANAGEMENT

(17) Indicator #6.1: Learning time is maximized

Item #34: *Teacher starts lesson on time.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #35: *Teacher makes sure that students are involved in learning activities until the end of the lesson.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #36: *Actions are taken to minimize disruption.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(18) Indicator #6.2: Clear rules are evident

Item #37: *There is clarity about when and how students can get help to do their work in class.*

5	**4**	**3**	**2**	**1**	**NA**
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

ISTOF Teacher Observation Protocol
(the EMT adapted version)

Item #38: *There is clarity about what options are available when the students finish their assignments.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

(19) Indicator #6.3: Misbehaviours and disruptions are effectively dealt with

Item #39: *The teacher corrects misbehaviour with measures that fit the seriousness of the misconduct (e.g., she does not overreact).*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

Item #40: *The teacher deals with misbehaviour and disruptions by referring to the established rules of the classroom.*

5	4	3	2	1	NA
Strongly Agree		*Neutral*		*Strongly Disagree*	*Not Applicable*

APPENDIX C: The Standardised Mathematics Test

March 2013

MATHEMATICS TEST

Calculator *not* allowed

Instructions:

- There are three types of questions: multiple-choice questions, short-answer questions, and extended answer questions:
  - To answer multiple-choice questions, please just fill in the appropriate letters A, B, C, or D.
  - To answer short-answer questions, please simply write down the answer.
  - To answer extended answer questions, please explain your ideas with words and/or diagrams.
- You have 40 minutes to complete the test.

First Name: ________________________

Last Name: ________________________

Class: ________________________

School: ________________________

June-July 2013

MATHEMATICS TEST

☺ Calculator *NOT* allowed ☺

Instructions:

- There are three types of questions: multiple-choice questions, short-answer questions, and extended answer questions:
  - To answer multiple-choice questions, please just fill in the appropriate letters A, B, C, or D.
  - To answer short-answer questions, please simply write down the answer.
  - To answer extended answer questions, please explain your ideas with words and/or diagrams.
- You have 40 minutes to complete the test.

First Name: ________________________

Last Name: ________________________

Class: ________________________

School: ________________________

Acknowledgement

This test is in the same version with the first test for the EMT project to measure pupils' learning gains over about ten school weeks. Test items are all derived from the item pool that IEA (International Association for the Evaluation of Educational Achievement) has released from TIMSS (Trends in International Mathematics and Science Study) 2003 for Grade-4 pupils (aged 9-10). TIMSS assessment framework has been systematically referred to during the process of item screening and organising so as to formulate an ideal booklet for the pre- and post-tests of this project. The authorship and copyright belong to IEA & TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, USA. Thanks to everyone who has designed and/or made these items accessible and reusable for researchers and practitioners.

1. □ represents the number of the magazines that Lina reads each week. Which of these represents the total number of magazines that Lina reads in 6 weeks?

Ⓐ 6+□ Ⓑ 6×□ Ⓒ □+6 Ⓓ (□+□)×6

Answer: ___________

2. 37 × ■ = 703.
What is the value of 37 × ■ + 6?

Answer: _____________

3. Here is a number pattern.

100, 1, 99, 2, 98, □, □, □

What three numbers should go in the boxes?

Ⓐ 3, 97, 4 Ⓑ 4, 97, 5 Ⓒ 97, 3, 96 Ⓓ 97, 4, 96

Answer: ______________

4.

4	11	6
		5
8	3	10

The rule for the table is that numbers in each row and column must add up to the same number. What number goes in the centre of the table?

Ⓐ 1 Ⓑ 2 Ⓒ 7 Ⓓ 12

Answer: _______________

1

5. The daily start times for showing a movie are listed below:

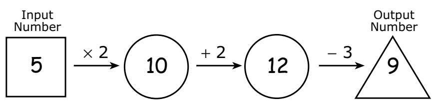
Show	Start Time
1st	2:00 p.m.
2nd	3:30 p.m.
3rd	5:00 p.m.
4th	?

If this pattern continues, what is the start time for the 4th show?

Ⓐ 5:30 p.m. Ⓑ 6:00 p.m. Ⓒ 6:30 p.m. Ⓓ 7:00 p.m.

Answer: ____________

6. A number machine takes a number and operates on it.
When the Input Number is 5, the Output Number is 9, as shown below.



When the Input Number is 7, which of these is the Output Number?

Ⓐ 11 Ⓑ 13 Ⓒ 14 Ⓓ 25

Answer: _____________

7. This chart shows temperature readings made at different times on four days.

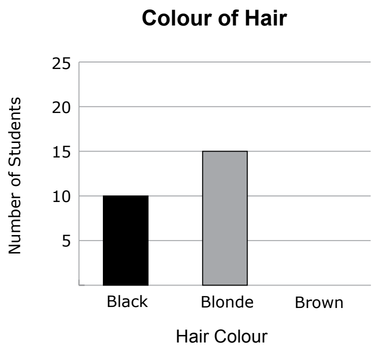
TEMPERATURES					
	6 a.m.	9 a.m.	Noon	3 p.m.	8 p.m.
Monday	15°	17°	20°	21°	19°
Tuesday	15°	15°	15°	10°	9°
Wednesday	8°	10°	14°	13°	15°
Thursday	8°	11°	14°	17°	20°

When was the highest temperature recorded?

Ⓐ Noon on Monday
Ⓑ 3 p.m. on Monday
Ⓒ Noon on Tuesday
Ⓓ 3 p.m. on Wednesday

Answer: _______________

8. In a class of 30 students, 10 have black hair, 15 have blonde hair, and the rest have brown hair. Complete the graph below to show the number of students with brown hair.



9. A shop owner decided to check how many pens, pencils, erasers, and rulers were sold on the day school opened. He made the tally chart below.

Pens	Pencils	Rubbers	Rulers
卌 卌 卌 卌 ||	卌 卌 卌 卌 卌 |	卌 卌 ||	卌 卌 卌 卌 ||||

How many more pencils than rulers were sold?

Answer: ____________

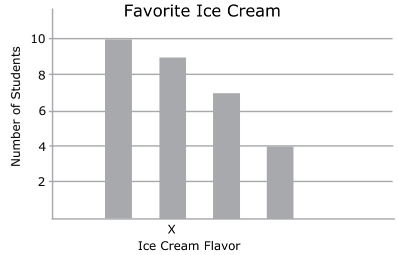
3

10.

Favorite Ice Cream	**Number of Students**
Butterscotch	IIII
Chocolate	𝍸 𝍸
Strawberry	𝍸 IIII
Vanilla	𝍸 II

A teacher asked 30 students in her class the flavour of their favourite ice cream. The table above shows how the teacher recorded the students' responses.

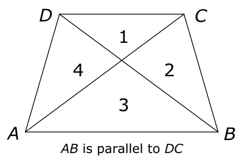
In the bar graph below, which ice cream flavour corresponds to the bar that is labelled X?



Ⓐ butterscotch Ⓑ chocolate Ⓒ strawberry Ⓓ vanilla

Answer: ____________

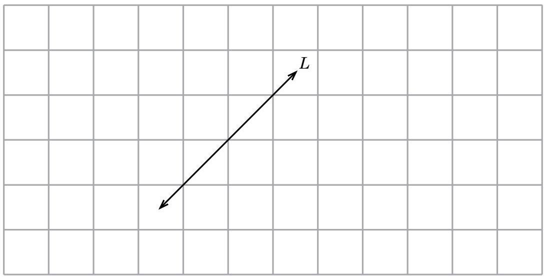
11.



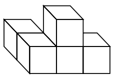
Two of the four triangles in the figure above are the same shape but different sizes. Shade in those two triangles.

4

12. On the grid, draw a line parallel to line L.

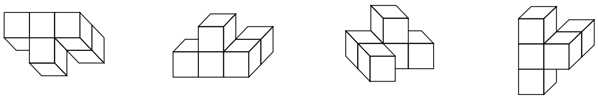


13. This figure will be turned to a different position.



Which of these could be the figure after it is turned?

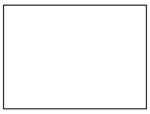
A B C D



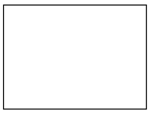
Answer: ____________

14.

A. Draw 1 straight line on this rectangle to divide it into 2 triangles.

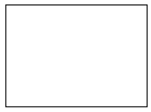


B. Draw 1 straight line on this rectangle to divide it into 2 rectangles.

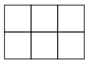


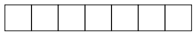
5

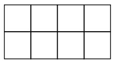
C. Draw 2 straight lines on this rectangle to divide it into 1 rectangle and 2 triangles.

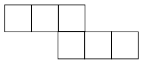


15. Which of these figures has the largest area?

Ⓐ 

Ⓑ 

Ⓒ 

Ⓓ 

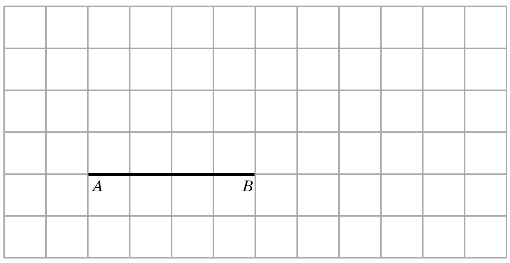
Answer: ______________

16. Simon wants to watch a film that is between $1\frac{1}{2}$ and 2 hours long. Which of the following films should he choose?

Ⓐ a 59-minute film
Ⓑ a 102-minute film
Ⓒ a 121-minute film
Ⓓ a 150-minute film

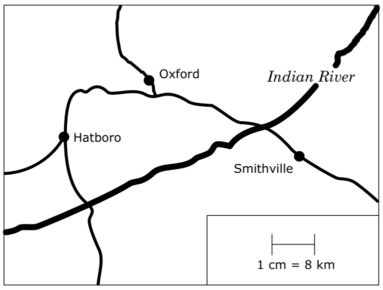
Answer: _______________

17.



Draw a triangle in the grid so that the line *AB* is the base of the triangle and the two new sides are the same length as each other.

18. One centimetre on the map represents 8 kilometres on the land.

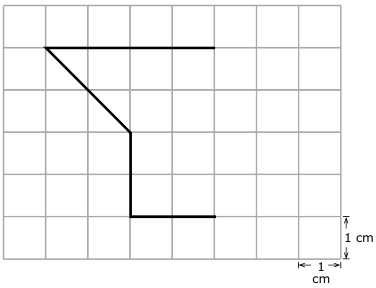


About how far apart are Oxford and Smithville on the land?

Ⓐ 4 km Ⓑ 16 km Ⓒ 35 km Ⓓ 50 km

Answer: ____________

19.



The squares in the grid above have areas of 1 square centimetre. Draw lines to complete the figure so that it has an area of 13 square centimetres.

7

20. Here is a calendar for December.

DECEMBER						
S	**M**	**T**	**W**	**T**	**F**	**S**
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Mary's birthday is on Thursday, 2nd December. She is going on a trip exactly 3 weeks later. On what date will she go on the trip?

Ⓐ 16th December Ⓑ 21st December
Ⓒ 23rd December Ⓓ 30th December

Answer: ___________

21. George practised soccer six days a week.
For 3 of the days he practised for 45 minutes each day.
For 3 of the days he practised for 20 minutes each day.

In hours and minutes, what is the total amount of time George practised on these six days?

Ⓐ 2 hours 20 minutes Ⓑ 2 hours 55 minutes
Ⓒ 3 hours 5 minutes Ⓓ 3 hours 15 minutes

Answer: __________

22. Jasmine made a stack of cubes of the same size. The stack had 5 layers and each layer had 10 cubes. What is the volume of the stack?

Ⓐ 5 cubes Ⓑ 15 cubes Ⓒ 30 cubes Ⓓ 50 cubes

Answer: ___________

23. Subtract:

$$\begin{array}{r} 4.03 \\ -\underline{1.15} \end{array}$$

Ⓐ 5.18 Ⓑ 4.45 Ⓒ 3.12 Ⓓ 2.88

Answer: ______________

24. Which of these means $\frac{7}{10}$?

Ⓐ 70 Ⓑ 7 Ⓒ 0.7 Ⓓ 0.07

Answer: ______________

25. There are 600 balls in a box, and $\frac{1}{3}$ of the balls are red.

How many red balls are in the box?

Answer: __________ red balls

26. For every soft drink bottle that Fred collected, Maria collected 3.
Fred collected a total of 9 soft drink bottles. How many did Maria collect?

Ⓐ 3 Ⓑ 12 Ⓒ 13 Ⓓ 27

Answer: ______________

27. What number equals 3 ones + 5 tens +4 hundreds + 60 thousands?

Ⓐ 6, 453 Ⓑ 60, 453 Ⓒ 64, 530

Ⓓ 354, 060 Ⓔ 604, 530

Answer: _____________

28. $204 \div 4 =$

Answer: ______________

29. Lisa is practising addition and substraction problems. What number should Lisa add to 142 to get 369?

Answer: ______________

30. Juanita wanted to use her calculator to add 1, 379 and 243. She entered 1, 279 + 243 by mistake. Which of these could she do to correct the mistake?

Ⓐ Add 100 Ⓑ Add 1 Ⓒ Subtract 1 Ⓓ Subtract 100

Answer: _______________

31. Mark's garden has 84 rows of cabbages. There are 57 cabbages in each row. Which of these gives the BEST way to estimate how many cabbages there are altogether?

Ⓐ $100 \times 50 = 5,000$
Ⓑ $90 \times 60 = 5,400$
Ⓒ $80 \times 60 = 4,800$
Ⓓ $80 \times 50 = 4,000$

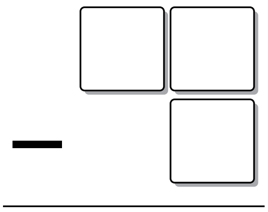
Answer: ________________

32. Each student needs 8 notebooks for school. How many notebooks are needed for 115 students?

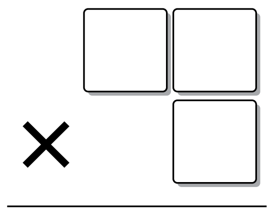
Answer: _________________

33. Using the number tiles, Joan and Herbert played a new game.
They placed the numbers to make the largest answer.

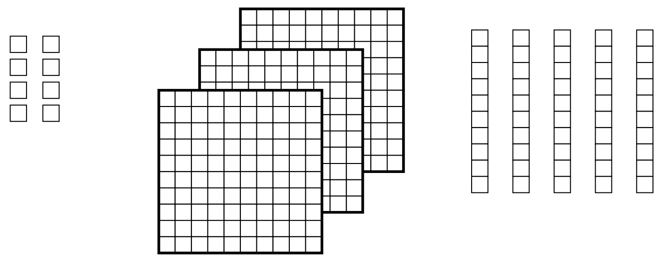
A. Use the tiles $\boxed{2}$, $\boxed{3}$, and $\boxed{7}$. Write the numbers on the tiles in the boxes below to make the largest answer when you subtract.



B. Use the tiles $\boxed{1}$, $\boxed{4}$, and $\boxed{5}$. Write the numbers on the tiles in the boxes below to make the largest answer when you multiply.



34. Each small square (☐) is equal to 1. There are 10 small squares in each strip. There are 100 small squares in each large square.



What number is shown?

Ⓐ 16 Ⓑ 358 Ⓒ 538 Ⓓ 835

Answer: ______________

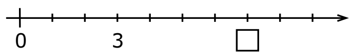
35. Which of these has the same value as 342?

Ⓐ 3, 000 + 400 + 2 Ⓑ 300 + 40+ 2

Ⓒ 30 + 4 + 2 Ⓓ 3 + 4 + 2

Answer: _______________

36.



On the number line above, what number goes in the box?

Number in ☐ = _______________

37. Get to 20 Number Game

Two children, Joan and Herbert, are learning to play a game "Get to 20." Here are the rules for the game.

GET TO 20
RULES

Pick Tiles: Each player draws three number tiles.

Add Tiles: Each player places the three tiles to make an addition problem with the sum total closest to 20.

For example, here are four ways a player who draws $\boxed{1}$, $\boxed{4}$, and $\boxed{5}$ could place the tiles:

$$\begin{array}{r} \boxed{5}\boxed{1} \\ +\ \boxed{4} \\ \hline 55 \end{array} \quad \textbf{or} \quad \begin{array}{r} \boxed{4}\boxed{5} \\ +\ \boxed{1} \\ \hline 46 \end{array} \quad \textbf{or} \quad \begin{array}{r} \boxed{1}\boxed{5} \\ +\ \boxed{4} \\ \hline 19 \end{array} \quad \textbf{or} \quad \begin{array}{r} \boxed{1} \\ +\ \boxed{5} \\ +\ \boxed{4} \\ \hline 10 \end{array}$$

This player should choose to show the addition problem $\begin{array}{r} 15 \\ +4 \\ \hline 19 \end{array}$ because 19 is the total closest to 20.

Herbert said, "If I pick $\boxed{1}$, $\boxed{4}$, and $\boxed{6}$, I can make 20 two different ways."

Show two ways Herbert could make 20 with $\boxed{1}$, $\boxed{4}$, and $\boxed{6}$.

First way:

Second way:

☺ End of Test ☺

APPENDIX D: Pupil Questionnaire

Pupil perception of mathematics teaching and learning

Acknowledgement:
This questionnaire has been developed on the basis of the questionnaire, Student Perception Survey for Elementary Student, constructed by the research team of the Measure of Effective Teaching project in the USA.

Dear Pupil,

Thank you for participating in this survey. While answering the questions, it is important that you think your experience in the mathematics lessons in this class.

Your teacher and head teacher will NOT look at your answer. Later they will be informed of how the students in your school responded, but not how you or any one individual student answered. Please answer what you really think and feel.

Zhenzhen Miao
Ph.D. student
Southampton Education School
University of Southampton
Southampton
SO17 1BJ
Email:Z.Miao@soton.ac.uk

☺**Your Name:** ______________________________

☺**Your Class:** ______________________________

☺**Your School:** ____________________________

	No, Never	Mostly Not	Maybe/ Some-times	Mostly Yes	Yes, Always
1. I like the ways we learn in maths lessons in this class.	☐	☐	☐	☐	☐
2. In maths lessons, you must pay attention all the time in order to keep up.	☐	☐	☐	☐	☐
3. In our maths lessons, mistakes are okay if you tried your best.	☐	☐	☐	☐	☐
4. My teacher is nice to me when I ask questions in maths.	☐	☐	☐	☐	☐
5. I have pushed myself hard to understand in maths lessons.	☐	☐	☐	☐	☐
6. If you don't understand something in maths, my teacher explains it another way.	☐	☐	☐	☐	☐
7. I like the way my teacher treats me when I need help in maths.	☐	☐	☐	☐	☐
8. My teacher pushes everybody to work hard in maths.	☐	☐	☐	☐	☐
9. In maths lessons, we learn to correct our mistakes.	☐	☐	☐	☐	☐
10. Our maths lessons stay busy and do not waste time.	☐	☐	☐	☐	☐
11. When he/she is teaching us maths, my teacher asks us whether we understand.	☐	☐	☐	☐	☐
12. In maths lessons, pupils behave so badly that it slows down our learning.	☐	☐	☐	☐	☐
13. Everybody knows what they should be doing and learning in maths lessons.	☐	☐	☐	☐	☐
14. My teacher explains difficult things in maths clearly.	☐	☐	☐	☐	☐
15. My teacher makes me want to go to university.	☐	☐	☐	☐	☐
16. My teacher in this class makes me feel that he/she really cares about me.	☐	☐	☐	☐	☐
17. My teacher doesn't let pupils give up when the work gets hard in maths.	☐	☐	☐	☐	☐

-1-

	No, Never	Mostly Not	Maybe/ Some-times	Mostly Yes	Yes, Always
18. If I need help with maths homework, there is someone at home who can help me.	☐	☐	☐	☐	☐
19. My teacher has several good ways to explain each topic that we cover in maths lessons.	☐	☐	☐	☐	☐
20. My teacher gives us time to explain our ideas in maths lessons.	☐	☐	☐	☐	☐
21. Homework helps me learn maths.	☐	☐	☐	☐	☐
22. In this class, our teacher accepts nothing less than our full effort.	☐	☐	☐	☐	☐
23. In maths lessons, my teacher knows when the class understands, and when we do not.	☐	☐	☐	☐	☐
24. We spend a lot of time practising for maths tests.	☐	☐	☐	☐	☐
25. In maths lessons, I take it easy and do not try very hard to do my best.	☐	☐	☐	☐	☐
26. In maths lessons, my teacher tells us what we are learning and why.	☐	☐	☐	☐	☐
27. Maths lessons in the class are often hard for the teacher to make clearer.	☐	☐	☐	☐	☐
28. Being in maths lessons makes me feel sad and angry.	☐	☐	☐	☐	☐
29. My after-school activities don’t leave enough time to finish my maths homework.	☐	☐	☐	☐	☐
30. We have interesting maths homework.	☐	☐	☐	☐	☐
31. In maths lessons, my teacher asks questions to be sure we are following along when he/she is teaching.	☐	☐	☐	☐	☐
32. When my teacher marks my maths work, he/she writes on my papers to help me understand how to do better.	☐	☐	☐	☐	☐
33. The teacher in this class encourages me to do my best in maths.	☐	☐	☐	☐	☐

-2-

	No, Never	Mostly Not	Maybe/ Some-times	Mostly Yes	Yes, Always
34. School work in maths is not very enjoyable. (Do you agree?)	☐	☐	☐	☐	☐
35. My teacher checks to make sure we understand what he/she is teaching us in maths lessons.	☐	☐	☐	☐	☐
36. In this class, doing your maths homework is very important. (Do you agree?)	☐	☐	☐	☐	☐
37. This class is a happy place for me to be.	☐	☐	☐	☐	☐
38. My teacher wants us to share our thoughts in maths.	☐	☐	☐	☐	☐
39. My teacher takes the time to summarise what maths knowledge we learn each day.	☐	☐	☐	☐	☐
40. Pupils get to decide how maths activities are done in this class.	☐	☐	☐	☐	☐
41. My teacher explains maths things in very orderly ways.	☐	☐	☐	☐	☐
42. When doing schoolwork in maths, I try to learn as much as I can and I don't worry about how long it takes.	☐	☐	☐	☐	☐
43. My teacher wants me to explain my answers in maths work – why I think what I think.	☐	☐	☐	☐	☐
44. Pupils don't share ideas in maths lessons; we mostly just listen to the teacher.	☐	☐	☐	☐	☐
45. I am happy with how well I have done in maths in this class.	☐	☐	☐	☐	☐
46. In maths lessons, we learn a lot almost every day.	☐	☐	☐	☐	☐
47. I understand what I am supposed to be learning in maths.	☐	☐	☐	☐	☐
48. I have done my best quality work in maths.	☐	☐	☐	☐	☐
49. At home, I don't have a quiet place where I can do homework. (Do you agree?)	☐	☐	☐	☐	☐
50. Pupils speak up and share their ideas about class work in maths.	☐	☐	☐	☐	☐

-3-

	No, Never	Mostly Not	Maybe/ Some-times	Mostly Yes	Yes, Always
51. This class is neat – everything has a place and things are easy to find.	☐	☐	☐	☐	☐
52. If I am sad or angry, my teacher helps me feel better.	☐	☐	☐	☐	☐
53. I have learned a lot this year about maths tests.	☐	☐	☐	☐	☐
54. In this class, pupils stop trying when the work in maths gets hard.	☐	☐	☐	☐	☐
55. My classmates behave the way my teacher wants them to.	☐	☐	☐	☐	☐
56. My teacher in this class does not know me very well yet.	☐	☐	☐	☐	☐
57. My teacher seems to know if something is bothering me.	☐	☐	☐	☐	☐
58. School work in maths is interesting.	☐	☐	☐	☐	☐
59. In maths lessons, getting right answer is very important.	☐	☐	☐	☐	☐
60. I think we get more maths homework in this class than children in other classes.	☐	☐	☐	☐	☐
61. For a new pupil, this class would be a good one to join.	☐	☐	☐	☐	☐
62. I read books at home almost everyday.	☐	☐	☐	☐	☐

63. When maths homework is assigned for this class, how much of it do you usually complete?
☐ Never Assigned ☐ None of it ☐ Some of it
☐ Most of it ☐ All ☐ All plus some extra

64. Outside of class, about how much time in a week do you usually spend doing homework in maths?
☐ No time ☐ Half an hour ☐ 1 hour ☐ 2 hours
☐ 3 to 4 hours ☐ 5 to 7 hours ☐ 8+ hours

65. During most weeks, how many days is there maths homework to do for this class? *(Select one choice.)*
☐ 1 day ☐ 2 days ☐ 3 days ☐ 4 days ☐ 5 days ☐ Never Assigned

-4-

66.	Are you a boy or a girl? ☐ Boy ☐ Girl
67.	Is there a computer at you house? ☐ No ☐ Yes, there is one ☐ Yes, there is more than one
68.	Does your family speak English at home? ☐ Yes, always ☐ Yes, sometimes ☐ No, never or almost never
69.	Counting yourself and all others, how many children live with you? ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 or more
70.	How many adults do you live with? ☐ One ☐ Two ☐ More than Two
71.	How many books do you think are in the room where you sleep? ☐ None ☐ Between 1 and 10 ☐ Between 11 and 24 ☐ More than 25
72.	What is you race/ethnicity? *(mark all that apply)* ☐ White British ☐ white (other) ☐ Indian ☐ Pakistani ☐ Mixed ☐ Black Caribbean ☐ Bangladeshi ☐ Chinese ☐ Other Asian (non Chinese) ☐ Black others ☐ Other
73.	When you are younger, what kind of marks (or grades) did you usually get in school? ☐ Very high ☐ High ☐ Good ☐ Some Good, Some Not ☐ Not Very Good
74.	How long have you been in this class? ☐ For less than two weeks ☐ For about three or four weeks ☐ Between one and two months ☐ More than two months

That's the end of the survey. Thanks a lot for your participation. ☺

-5-

APPENDIX E: Teacher Questionnaire

Teacher Questionnaire on the Teaching of Mathematics

Acknowledgement
This questionnaire is developed on the basis of the Teacher Questionnaire in the Teaching and Learning International Survey (TALIS 2008) conducted by the Organisation for Economic Cooperation and Development (OECD) in 2008 with certain items revised or deleted in order to suit the research design of this study.

Dear Teacher,

Thank you very much for your participation in this survey. Please answer what you really think and feel. Anyone other than the researcher will not read your answers. Your school leader will only be informed of how all participants of the region responded, but not how you and any individual teacher answered. Your name, your class name and school name will be anonymised in the final report and in any publication of this study.

Zhenzhen Miao
Ph.D. student
Southampton Education School
University of Southampton
Southampton
SO17 1BJ
Email:Z.Miao@soton.ac.uk

☺**Name**: ____________________________

☺**School**: ___________________________

☺**Date**: ____________________________

Your participation and opinions are sincerely appreciated!

Background information

These questions are about you, your education and the time you have spent in teaching. In responding to the questions, please mark the appropriate box.

1. What is your gender?

Female ☐ Male ☐

2. How old are you?

Under 25	25-29	30-39	40-49	50-59	60+
☐	☐	☐	☐	☐	☐

3. What is your employment status as a teacher?

Part-time employment is where the contracted hours of work represent less than 90 per cent of the normal or statutory number of hours of work for a full-time employee over a complete school year. Please consider your employment status for all of your teaching jobs combined.

☐ Full-time

☐ Part-time (50-90% full-time hours)

☐ Part-time (less than 50% of full-time hours)

4. Do you work as a primary teacher at another school as well as this school?

☐ Yes

☐ No → Please go to question 6.

5. If 'Yes' in the previous questions, please indicate in how many other schools you work as a primary teacher.

Please write in a number.

☐☐ Schools

1

6. What is your employment status as a teacher at this school?

Please do not consider the probationary period of a contract as a separate contract.

☐ Permanent employment (an on-going contract with no fixed end-point before the age of retirement)

☐ Fixed term contract for a period of more than 1 school-year

☐ Fixed-term contract for a period of 1 school-year or less

7. What is the highest level of formal education that you have completed?

Please mark one choice.

☐ Bachelor degree

☐ Masters degree

☐ Doctoral degree

8. In a typical school week, estimate the number of (60 minutes) hours you spend on the following for this school.

This question concerns your work for this school only. Please do not include the work you do for other schools.

Please write a number in each row and round to the nearest hour in your responses. Write 0 (zero) if none.

a) ☐☐ Teaching of pupils in school (either whole class, in groups or individually)

b) ☐☐ Planning or preparation of lessons either in school or out of school (including marking of pupil work)

c) ☐☐ Administrative duties either in school or out of school (including school administrative duties, paperwork and other clerical duties you undertake in your job as a teacher)

d) ☐☐ Other (please specify):____________________________

2

9. How long have you been working as a teacher?

Where possible exclude extended periods of absence (e.g. career breaks).

This is my first year	1-2 years	3-5 years	6-10 years	11-15 years	16-20 years	More than 20 years
☐	☐	☐	☐	☐	☐	☐

10. How long have you been working as a teacher <u>at this school</u>?

Where possible exclude extended periods of absence (e.g. career breaks).

This is my first year	1-2 years	3-5 years	6-10 years	11-15 years	16-20 years	More than 20 years
☐	☐	☐	☐	☐	☐	☐

Please turn to the next page for a new section.

Professional Development

*In this survey, **professional development** is defined as activities that develop an individual's skills, knowledge, expertise and other characteristics as a teacher.*

*Please only consider professional development you have taken **after** your initial teacher training/education.*

11. During the last 18 months, did you participate in any of the following kinds of professional development activities, and what was the impact of these activities on your development as a teacher?

For each question below, please mark one choice in part (A). If you answer 'Yes' in part (A) then please mark one choice in part (B) to indicate how much impact it had upon your development as a teacher.

	(A) Participation		(B) Impact			
	Yes	No	No impact	A small impact	A moderate impact	A large impact
a) Courses/workshops (e.g. on subject matter or methods and/or other education-related topics) …..	☐	☐	☐	☐	☐	☐
b) Education conferences or seminars (where teachers and/or researchers present their research results and discuss educational problems) ………………………………..	☐	☐	☐	☐	☐	☐
c) Qualification programme (e.g. a degree programme) ……………………	☐	☐	☐	☐	☐	☐
d) Observation visits to other schools ………………………………………………	☐	☐	☐	☐	☐	☐
e) Participation in a network of teachers formed specifically for the professional development of teachers ………………………………..	☐	☐	☐	☐	☐	☐
f) Individual or collaborative research on a topic of interest to you professionally ………………………..	☐	☐	☐	☐	☐	☐
g) Mentoring and/or peer observation and coaching, as part of a formal school arrangement …..	☐	☐	☐	☐	☐	☐

4

12. In all, how many days of professional development did you attend during the last 18 months?

Please round to whole days. Write 0 (zero) if none.

Days

If you answered '0' (zero) → Please go to question 17.

13. Of these, how many days were compulsory for you to attend as part of your job as a teacher?

Please round to whole days. Write 0 (zero) if none.

Days

14. For the professional development in which you participated in the last 18 months, how much did you personally have to pay for?

Please mark one choice.

None	Some	All
☐	☐	☐

15. For the professional development in which you participated in the last 18 months, did you receive scheduled time for undertaking the professional development that took place <u>during</u> regular work hours?

Please mark one choice.

☐ Yes

☐ No

☐ Did not take place during regular work hours

16. For the professional development in which you participated in the last 18 months, did you receive a salary supplement for undertaking the professional development activities that took place outside regular work hours?

Please mark one choice.

☐ Yes

☐ No

☐ Did not take place outside of regular work hours

17. Thinking about less formal professional development, during the last 18 months, did you participate in any of the following activities, and what was the impact of these activities on your development as a teacher?

For each question below, please mark one choice in part (A). If you answer 'Yes' in part (A) then please mark one choice in part (B) to indicate how much impact it had upon your development as a teacher.

	(A) Participation		(B) Impact			
	Yes	No	No impact	A small impact	A moderate impact	A large impact
a) Reading professional literature (e.g. journals, evidence-based papers, thesis papers)	☐	☐	☐	☐	☐	☐
b) Engaging in informal dialogue with your colleagues on how to improve your teaching	☐	☐	☐	☐	☐	☐

18. Thinking of your own professional development needs, please indicate the extent to which you have such needs in each of the areas listed.

Please mark one choice in each row.

	No need at all	Low level of need	Moderate level of need	High level of need
a) Content and performance standards in my main subject field(s)	☐	☐	☐	☐
b) Pupil assessment practices	☐	☐	☐	☐
c) Classroom management	☐	☐	☐	☐
d) Knowledge and understanding of my main subject field(s)	☐	☐	☐	☐
e) Knowledge and understanding of instructional practices (knowledge mediation) in my main subject field(s)	☐	☐	☐	☐

6

	No need at all	Low level of need	Moderate level of need	High level of need
f) ICT skills for teaching …………………………………	☐	☐	☐	☐
g) Teaching pupils with special learning needs …………………………………………………………	☐	☐	☐	☐
h) Pupil discipline and behaviour problems …	☐	☐	☐	☐
i) School management and administration ……..	☐	☐	☐	☐
j) Teaching in multicultural settings ……………….	☐	☐	☐	☐
k) Pupil counselling …………………………………	☐	☐	☐	☐

19. In the last 18 months, did you want to participate in more professional development than you did?

☐ Yes

☐ No → **Please go to question 21.**

20. If 'Yes' in the previous question, which of the following reasons best explain what prevented you from participating in more professional development than you did?

Please mark as many choices as appropriate.

☐ I did not have the pre-requisites (e.g. qualifications, experiences, seniority).

☐ Professional development was too expensive /I could not afford it.

☐ There was a lack of employer support.

☐ Professional development conflicted with my work schedule.

☐ I didn't have time because of family responsibilities.

☐ There was no suitable professional development offered.

☐ Other (please specify): ____

Please turn to the next page for a new section.

Teacher Appraisal and Feedback

We would like to ask you about the appraisal (defined below) of your work as a teacher and the feedback (defined below) you receive about your work in this school.

*In this survey, **Appraisal** is defined as when a teacher's work is reviewed by the head teacher, an external inspector or by his or her colleagues. This appraisal can be conducted in a range of ways from a more formal, objective approach (e.g. as part of a formal performance management system, involving set procedures and criteria) to the more informal, more subjective approach (e.g. through informal discussions with the teacher).*

*In this survey, **Feedback** is defined as the reporting of the results of a review of your work (however formal or informal that review has been) back to the teacher, often with the purpose of noting good performance or identifying areas for development. Again, the feedback may be provided formally (e.g. through a written report) or informally (e.g. through discussions with the teachers).*

21. From the following people, how often have you received appraisal and/or feedback about your work as a teacher in this school?

Please mark one choice in each row.

	Never	Less than once every two years	Once every two years	Once per year	Twice per year	3 or more times per year	Monthly	More than once per month
a) Principal ………………	☐	☐	☐	☐	☐	☐	☐	☐
b) Other teachers or members of the school management team …………………	☐	☐	☐	☐	☐	☐	☐	☐
c) External individual or body (e.g. external inspector) ……………………………	☐	☐	☐	☐	☐	☐	☐	☐

If you answered 'Never' for all of the above (a, b, and c) → Please go to question 28.

8

22. In your opinion, how important were the following aspects considered to be when you received this appraisal and/or feedback?

Please mark one choice in each row.

	I do not know if it was considered	Not considered at all	Considered with low importance	Considered with moderate importance	Considered with high importance
a) Pupil test scores	☐	☐	☐	☐	☐
b) Retention and pass rates of pupils	☐	☐	☐	☐	☐
c) Other pupil learning outcomes	☐	☐	☐	☐	☐
d) Pupil feedback on my teaching	☐	☐	☐	☐	☐
e) Feedback from parents	☐	☐	☐	☐	☐
f) How well I work with the head teacher and my colleagues	☐	☐	☐	☐	☐
g) Direct appraisal of my classroom teaching	☐	☐	☐	☐	☐
h) Innovative teaching practices	☐	☐	☐	☐	☐
i) Relations with pupils	☐	☐	☐	☐	☐
j) Professional development I have undertaken...	☐	☐	☐	☐	☐
k) Classroom management	☐	☐	☐	☐	☐
l) Knowledge and understanding of my main subject field(s)	☐	☐	☐	☐	☐
m) Knowledge and understanding of instructional practices (knowledge mediation) in my main subject field(s) ...	☐	☐	☐	☐	☐
n) Teaching pupils with special learning needs	☐	☐	☐	☐	☐
o) Pupil discipline and behaviour	☐	☐	☐	☐	☐
p) Teaching in a multicultural settings	☐	☐	☐	☐	☐
q) Extra-curricular activities with pupils (e.g. school plays and performances, sporting activities)	☐	☐	☐	☐	☐

9

	I do not know if it was considered	Not considered at all	Considered with low importance	Considered with moderate importance	Considered with high importance
r) Other (please specify below) …………………	☐	☐	☐	☐	☐

____

23. Concerning the appraisal and/or feedback you have received at this school, to what extent have they directly led to any of the following?

Please mark one choice in each row.

	No change	A small change	A moderate change	A large change
a) A change in salary …………………………………….	☐	☐	☐	☐
b) A financial bonus or another kind of monetary reward ………………………………………	☐	☐	☐	☐
c) Opportunities for professional development activities ………………………………………………	☐	☐	☐	☐
d) A change in the likelihood of career advancement …………………………………………	☐	☐	☐	☐
e) Public recognition from the head teacher and/or your colleagues ……………………………	☐	☐	☐	☐
f) Changes in your work responsibilities that make the job more attractive ……………………	☐	☐	☐	☐
g) Role in school development initiatives (e.g. curriculum development group, development of school objectives) ……………	☐	☐	☐	☐

24. Concerning the appraisal and/or feedback you have received at this school, to what extent have they directly led to or involved changes in any of the following?

Please mark one choice in each row.

	No change	A small change	A moderate change	A large change
a) Your classroom management practices	☐	☐	☐	☐
b) Your knowledge and understanding of your main subject field(s)	☐	☐	☐	☐
c) Your knowledge and understanding of instructional practices (knowledge mediation) in your main subject field(s)	☐	☐	☐	☐
d) A development or training plan to improve your teaching	☐	☐	☐	☐
e) Your teaching of pupils with special learning needs	☐	☐	☐	☐
f) Your handling of pupil discipline and behaviour problems	☐	☐	☐	☐
g) Your teaching of pupils in a multicultural setting	☐	☐	☐	☐
h) The emphasis you place upon improving pupil test scores in your teaching	☐	☐	☐	☐

25. How would you describe the appraisal and/or feedback you received?

Please mark one choice in each row.

	Yes	No
a) The appraisal and/or feedback contained a judgment about the quality of my work.	☐	☐
b) The appraisal and/or feedback suggestions for improving certain aspects of my work	☐	☐

11

26. Regarding the appraisal and/or feedback you received at this school, to what extent do you agree or disagree with the following statements?

Please mark one choice in each row.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a) I think the appraisal of my work and/or feedback received was a fair assessment of my work as a teacher in this school. ………………………	☐	☐	☐	☐
b) I think the appraisal of my work and/or feedback received was helpful in the development of my work as a teacher in this school. ……………………………………………………………	☐	☐	☐	☐

27. Concerning the appraisal and/or feedback you have received at this school, to what extent have they directly led to any of the following?

Please mark one choice in each row.

	A large decrease	A small decrease	No change	A small increase	A large increase
a) Changes in your job satisfaction …………….	☐	☐	☐	☐	☐
b) Changes in your job security …………………..	☐	☐	☐	☐	☐

28. We would like to ask you about appraisal and/or feedback to teachers in this school more generally. To what extent do you agree or disagree with the following statements?

Please mark one choice in each row.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a) In my opinion, in this school the head teacher takes steps to alter the monetary rewards of a persistently underperforming teacher.	☐	☐	☐	☐
b) In my opinion, in this school the sustained poor performance of a teacher would be tolerated by the rest of the staff.	☐	☐	☐	☐
c) In this school, teachers will be dismissed because of sustained poor performance.	☐	☐	☐	☐
d) In my opinion, in this school the head teacher uses effective methods to determine whether teachers are performing well or badly.	☐	☐	☐	☐
e) In my opinion, in this school a development or training plan is established for teachers to improve their work as a teacher.	☐	☐	☐	☐
f) In my opinion, the most effective teachers in this school receive greatest monetary or nonmonetary rewards.	☐	☐	☐	☐
g) If I improve the quality of my teaching at this school, I will receive increased monetary or nonmonetary rewards.	☐	☐	☐	☐
h) If I'm more innovative in my teaching at this school, I will receive increased monetary or nonmonetary rewards.	☐	☐	☐	☐
i) In my opinion, in this school the review of teachers' work is largely done to fulfil administrative requirements.	☐	☐	☐	☐
j) In my opinion, in this school the review of teachers' work has little impact upon the way teachers teach in the classroom.	☐	☐	☐	☐

Please turn to the next page for a new section.

13

Teaching Practices, Beliefs and Attitudes

29. We would like to ask about your personal beliefs on teaching and learning. Please indicate how much you disagree or agree with each of the following statements.
Please mark one choice in each row.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a) Effective/good teachers demonstrate the correct way to solve a problem.	☐	☐	☐	☐
b) When referring to a "poor performance", I mean a performance that lies below the previous achievement level of the pupil.	☐	☐	☐	☐
c) It is better when the teacher – not the pupil – decides what activities are to be done.	☐	☐	☐	☐
d) My role as a teacher is to facilitate pupils' own inquiry.	☐	☐	☐	☐
e) Teachers know a lot more than pupils; they shouldn't let pupils develop answers that may be incorrect when they can just explain the answers directly.	☐	☐	☐	☐
f) Pupils learn best by finding solutions to problems on their own.	☐	☐	☐	☐
g) Instruction should be built around problems with clear, correct answers, and around ideas that most pupils can grasp quickly.	☐	☐	☐	☐
h) How much pupils learn depends on how much background knowledge they have – that is why teaching facts is so necessary.	☐	☐	☐	☐
i) Pupils should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved.	☐	☐	☐	☐
j) When referring to a "good performance", I mean a performance that lies above the previous achievement level of the pupil.	☐	☐	☐	☐
k) A quiet classroom is generally needed for effective learning.	☐	☐	☐	☐
l) Thinking and reasoning processes are more important than specific curriculum content.	☐	☐	☐	☐

30. How often do you do the following in this school?

Please mark one choice in each row.

	Never	Less than once per year	Once per year	3-4 times per year	Monthly	Weekly
a) Attend staff meetings to discuss the vision and mission of the school	☐	☐	☐	☐	☐	☐
b) Develop a school curriculum or part of it	☐	☐	☐	☐	☐	☐
c) Discuss and decide on the selection of instructional media (e.g. textbooks, exercise books)	☐	☐	☐	☐	☐	☐
d) Exchange teaching materials with colleagues	☐	☐	☐	☐	☐	☐
e) Attend team conferences for the age group I teach	☐	☐	☐	☐	☐	☐
f) Ensure common standards in evaluations for assessing pupil progress	☐	☐	☐	☐	☐	☐
g) Engage in discussion about the learning development of specific pupils	☐	☐	☐	☐	☐	☐
h) Teach jointly as a team in the same class	☐	☐	☐	☐	☐	☐
i) Take part in professional learning activities (e.g. team supervision)	☐	☐	☐	☐	☐	☐
j) Observe other teachers' classes and provide feedback	☐	☐	☐	☐	☐	☐
k) Engage in joint activities across different classes and age groups (e.g. projects)	☐	☐	☐	☐	☐	☐
l) Discuss and coordinate homework practice across subjects	☐	☐	☐	☐	☐	☐

15

31. How strongly do you agree or disagree with the following statements …

Please mark one choice in each row.

… about yourself as a teacher in this school?	Strongly Disagree	Disagree	Agree	Strongly Agree
a) All in all, I am satisfied with my job. ………………………	☐	☐	☐	☐
b) I feel that I am making a significant educational difference in the lives of my pupils. …………………	☐	☐	☐	☐
c) If I try really hard, I can make progress with even the most difficult and unmotivated pupils. ………	☐	☐	☐	☐
d) I am successful with the pupils in my class. ……..	☐	☐	☐	☐
e) I usually know how to get through to pupils. ….	☐	☐	☐	☐
f) Teachers in this local community are well respected. ……………………………………………………………	☐	☐	☐	☐

… about what happens in this school?	Strongly Disagree	Disagree	Agree	Strongly Agree
g) In this school, teachers and pupils usually get on well with each other. ……………………………………………	☐	☐	☐	☐
h) Most teachers in this school believe that pupils' well-being is important. ……………………………………	☐	☐	☐	☐
i) Most teachers in this school are interested in what pupils have to say. ……………………………………………	☐	☐	☐	☐
j) If a pupil from this school needs extra assistance, the school provides it. ………………………..	☐	☐	☐	☐

32. Below you can find statements about the management of your school. Please indicate your perceptions of the frequency with which these activities took place during the current school year.

Please mark one choice in each row.

	Never	Seldom	Quite often	Very often
a) In meetings, the head teacher discusses educational goals with teachers.	☐	☐	☐	☐
b) The head teacher ensures that teachers work according to the school's educational goals.	☐	☐	☐	☐
c) The head teacher or someone else in the management team observes teaching in classes.	☐	☐	☐	☐
d) The head teacher gives teachers suggestions as to how they can improve their teaching.	☐	☐	☐	☐
e) When a teacher has problems in his/her classroom, the head teacher takes the initiative to discuss the matter.	☐	☐	☐	☐
f) The head teacher ensures that teachers are informed about possibilities for updating their knowledge and skills.	☐	☐	☐	☐
g) The head teacher compliments teachers for special effort or accomplishments.	☐	☐	☐	☐
h) In this school, the head teacher and teachers work on a school development plan.	☐	☐	☐	☐
i) The head teacher defines goals to be accomplished by the staff of this school.	☐	☐	☐	☐
j) The head teacher ensures that a task-oriented atmosphere is fostered in this school.	☐	☐	☐	☐
k) In this school, the head teacher and teachers act to ensure that the education quality issues are a collective responsibility.	☐	☐	☐	☐

17

33. We would like to ask you about the main subjects that you teach in this school in this school year.

Please indicate the subjects that you teach in this school (indicate only those that individually account for at least 20% of your teaching time in this school). The exact name of your subjects may not appear in the list below each category. If it does not, please mark the category you think best fits the subject.

	Yes	No
a) Reading, writing and literature ……… *Includes reading and writing (and literature) in the mother tongue, reading and writing (and literature) in the language of instruction, reading and writing in the tongue of the country (region) as a second language (for non-natives), language studies, public speaking, literature.*	☐	☐
b) Mathematics ……… *Includes mathematics, mathematics with statistics, geometry, algebra etc.*	☐	☐
c) Science ……… *Includes science, physics, physical science, chemistry, biology, human biology, environmental science, agriculture/horticulture/forestry.*	☐	☐
d) Modern foreign languages ……… *Includes languages different from the language of instruction.*	☐	☐
e) Technology ……… *Includes orientation in technology, including information technology, computer studies, construction/surveying, electronics, graphics and design, keyboard skills, word processing, workshop technology / design technology.*	☐	☐
f) Arts ……… *Includes arts, music, visual arts, practical art, drama, performance music, photography, drawing, creative handicraft, creative needlework.*	☐	☐
g) Physical education ……… *Includes physical education, gymnastics, dance, health.*	☐	☐
h) Religion and/or ethics ……… *Includes religion, history of religions, religion culture, ethics.*	☐	☐
i) Other (please specify below) ………	☐	☐

____

Please turn to the next page for a new section.

18

Your Teaching in the Observed Class

The following questions ask you about the class that you teach and has been observed in the study.

34. On average throughout the year how many pupils are in the class?

Please write a number.

Number of pupils

35. How would describe the ability of pupils in this class?

Please mark one choice in each row.

	Much lower than average ability	Slightly lower than average ability	Average ability	Slightly higher than average ability	Much higher than average ability
a) Compared to other pupils in the same grade/year level **in this school**?	☐	☐	☐	☐	☐
b) Compared to other pupils in the same grade/year level more generally	☐	☐	☐	☐	☐

36. For this class, please estimate a broad percentage of pupils who have the following characteristics.

It is acceptable to base your replies on rough estimates.
Please mark one choice in each row.

	Less than 10%	10% or more but less than 20%	20% or more but less than 40%	40% or more but less than 60%	60% or more
a) Pupils whose <first language> is different from the language(s) of instruction or a dialect of this/these	☐	☐	☐	☐	☐
b) Pupils who have at least one parent/guardian who has completed secondary education or higher	☐	☐	☐	☐	☐
c) Pupils who have at least one parent/guardian who has completed Bachelor degree or higher	☐	☐	☐	☐	☐

19

37. For <this class>, what percentage of <class> time is typically spent on each of the following activities?

Write a percentage for each activity. Write 0 (zero) if none.
Please ensure that responses add up to 100%.

a) ☐☐☐ % Administrative tasks (e.g. recording attendance, handing out school information/forms)

b) ☐☐☐ % Keeping order in the classroom (maintaining discipline)

c) ☐☐☐ % Actual teaching and learning

1 0 0 % Total

38. How often do you do the following activities happen in <maths lessons> in this class throughout the school year?

Please note that not all questions in this section are fully adapted to all sorts of teachers. Therefore, please just answer as best you can.
Please mark one choice in each row.

	Never or hardly never	In about one-quarter of maths lessons	In about one-half of maths lessons	In about three quarters of maths lessons	In almost every maths lesson
a) I present new topics to the class (lecture style presentation).	☐	☐	☐	☐	☐
b) I explicitly state learning goals.	☐	☐	☐	☐	☐
c) I review with pupils the homework they have prepared.	☐	☐	☐	☐	☐
d) Pupils work in small groups to come up with a joint solution to a problem or task.	☐	☐	☐	☐	☐

20

	Never or hardly never	In about one-quarter of maths lessons	In about one-half of maths lessons	In about three quarters of maths lessons	In almost every maths lesson
e) I give different work to the pupils that have difficulties learning and/or to those who can advance faster.................	☐	☐	☐	☐	☐
f) I ask my pupils to suggest or help plan classroom activities or topics.	☐	☐	☐	☐	☐
g) I ask my pupils to remember every step in a procedure.	☐	☐	☐	☐	☐
h) At the beginning of the lesson I present a short summary of the previous lesson.	☐	☐	☐	☐	☐
i) I check my pupils' exercise books.	☐	☐	☐	☐	☐
j) Pupils work on projects that require at least one week to complete.	☐	☐	☐	☐	☐
k) I work with individual pupils.	☐	☐	☐	☐	☐
l) Pupils evaluate or reflect upon their own work.	☐	☐	☐	☐	☐
m) I check, by asking questions, whether or not the subject matter has been understood.	☐	☐	☐	☐	☐
n) Pupils work in groups based upon their abilities.	☐	☐	☐	☐	☐
o) Pupils make a product that will be used by someone else.	☐	☐	☐	☐	☐
p) I administer a test or quiz to assess pupil learning.	☐	☐	☐	☐	☐
q) I ask my pupils to write an essay in which they are expected to explain their thinking or reasoning at some length. ...	☐	☐	☐	☐	☐

21

	Never or hardly never	In about one-quarter of maths lessons	In about one-half of maths lessons	In about three quarters of maths lessons	In almost every maths lesson
r) Pupils work individually with textbook or work sheets to practice newly taught subject matter. ………………	☐	☐	☐	☐	☐
s) Pupils hold a debate and argue for a particular point of view which might not be their own. …………..…………………………	☐	☐	☐	☐	☐

39. How strongly do you agree or disagree the following statements about <this class>?

Please mark one choice in each row.

	Strongly Disagree	Disagree	Agree	Strongly Agree
a) When the lesson begins, I have to wait for quite a long time for pupils to <quieten down>. ………..	☐	☐	☐	☐
b) Pupils in this class take care to create a pleasant learning atmosphere. ………………………………………	☐	☐	☐	☐
c) I lose quite a lot of time because of pupils interrupting the lesson. …………………………………….	☐	☐	☐	☐
d) There is much noise in this classroom. ………………..	☐	☐	☐	☐

22

☺ **You've completed the questionnaire. Many Thanks & Best Wishes to you!**

23

APPENDIX F: EMT Data Collection Timetable

	Obs	T1	T2	TQ	PQ	itv	FG
CN1	10/12/12	28/03/13	07/06/13	25/12/12	10/12/12	11/12/12	25/12/12
CN2	10/12/12	28/03/13	07/06/13	25/12/12	10/12/12	11/12/12	25/12/12
CN3	10/12/12	28/03/13	07/06/13	25/12/12	10/12/12	11/12/12	25/12/12
CN4	13/12/12	28/03/13	07/06/13	25/12/12	13/12/12	25/12/12	25/12/12
CN5	13/12/12	28/03/13	07/06/13	25/12/12	13/12/12	25/12/12	25/12/12
CN6	13/12/12	28/03/13	07/06/13	25/12/12	13/12/12	25/12/12	25/12/12
CN7	20/12/12	28/03/13	07/06/13	20/12/12	20/12/12	20/12/12	26/12/12
CN8	24/12/12	28/03/13	07/06/13	24/12/12	24/12/12	24/12/12	26/12/12
CN9	24/12/12	28/03/13	07/06/13	24/12/12	24/12/12	24/12/12	26/12/12
EN1	28/11/12	28/03/13	27/06/13	26/02/13	26/02/13	26/02/13	06/06/13
EN2	29/11/12	28/03/13	27/06/13	26/02/13	26/02/13	26/02/13	06/06/13
EN3	12/03/13	28/03/13	28/06/13	14/05/13	14/05/13	12/03/13	06/06/13
EN4	09/05/13	28/03/13	28/06/13	14/05/13	14/05/13	09/05/13	06/06/13
EN5	14/05/13	28/03/13	28/06/13	14/05/13	14/05/13	14/05/13	06/06/13
EN6	29/04/13	26/04/13	15/07/13	10/05/13	10/05/13	01/05/13	22/05/13
EN7	30/04/13	26/04/13	15/07/13	10/05/13	10/05/13	01/05/13	22/05/13
EN8	01/05/13	26/04/13	15/07/13	10/05/13	10/05/13	01/05/13	22/05/13
EN9	02/05/13	26/04/13	15/07/13	10/05/13	10/05/13	08/05/13	22/05/13
EN10	07/05/13	26/04/13	15/07/13	10/05/13	10/05/13	08/05/13	22/05/13

Note:
All dates in the British format DD/MM/YY. Obs = observation; T1 = test 1; T2 = test 2; TQ = teacher questionnaire; PQ = pupil questionnaire; itv = interview; FG = focus group.

APPENDIX G: Publishing & Disseminating Research

Article

- Miao, Z., Reynolds, D., Harris, A., & Jones, M. (2015). Comparing performance: a cross-national investigation into the teaching of mathematics in primary classrooms in England and China. *Asia Pacific Journal of Education, 35*(3), 392-403. doi:10.1080/02188791.2015.1056593

Conference

- Miao, Z. & Reynolds, D. (2015). Uncovering classroom-level factors for the performance gap in mathematics between England and China: a synthesis of results and findings from an international teacher effectiveness study. In, *The Annual Conference of British Educational Research Association (BERA)*, Belfast, GB, 15 - 17 September 2015.

- Miao, Z. & Reynolds, D. (2014). The effectiveness of mathematics teaching: a cross-national investigation in primary schools in England and China. In, *The Annual Conference of British Educational Research Association (BERA)*, London, GB, 23 - 25 September 2014. (press released)

- Miao, Z. & Reynolds, D. (2014). A comparative study on the effectiveness of mathematics teaching in England and China. In, *The European Association for Research on Learning and Instruction (EARLI) SIG 18 Conference*, Southampton, GB, 27 - 29 August 2014.