

**EXAMINING THE ALIGNMENT OF GRADING POLICIES
IN THE CHINESE EDUCATION SYSTEM**

by

Cheng Zhou

A project submitted to the Faculty of Education
in conformity with the requirements for
the degree of Master of Education

Queen's University
Kingston, Ontario, Canada
(October, 2015)

Copyright © Cheng Zhou, 2015

Abstract

A *grade* is a prevalent symbol used to communicate student performance over a period of learning. *Grading* is the process of assigning a *grade*, which can be affected by a variety of determinant factors (e.g., achievement and non-achievement factors), scales, reporting formats, and other socio-personal factors (e.g., subjects, schools, teachers' personal beliefs, and teachers' values). Grading policies, standards, and documents are published to establish and maintain consistencies for teachers' grading practices. However, grading policies are evident across multiple levels within an educational system (e.g., governmental, state/provincial, school district, school family, school, grade level), which can contribute to diverse conceptions of grading causing potential discrepancies related to grade determination, interpretation, and use.

This study examines the alignment of grading policies at three educational levels—national, provincial and school—within the Chinese education system. This study responds to the following questions: (a) how do national, provincial, and school level grading policies describe the purpose and process of grading, and (b) what is the degree of alignment across grading policies from national, provincial, and school levels. Results from this study find that there are points of alignment and misalignment related to the purpose and process of grading across system levels. Findings from this study may contribute to knowledge of grades and grading in the Chinese context, and may also serve as a reference for future grading policy and grading related studies.

Acknowledgements

“Everything you see exists together in a delicate balance.”

--The Lion King

Started my MEd study two years ago, I was excited. Came to a complete new country, a new city and immersed myself into a different educational, cultural and social environment was a big change. Too many challenges once made me anxious. However, with the help and support from professors, colleagues, friends and parents, I gradually find my own pace and walk all the way to the graduation.

A wise professor once commented that graduate study was a life changing experience, which I thought was a motivated but a little bit exaggerated description at that time. But now, I think I will say “YES”!

For me, the changing point was all about having a better understanding of the word “balance”. Not only from the changes of my own life, the study and living experiences in Canada, but also from the grading project I did. I realised that the word “balance” gradually became the key word of this journey. In life, I tried to balance different roles, duties and tasks that I had. It was about commitment, management, communication, cooperation and mutual understanding. In the writing process, I tried to balance the inner voice of my own, the evidence from previous empirical studies as well as the policy documents that I collected. And for the grading project itself, the ultimate goal was also to have a balanced grading system, which would provide better, more reliable and more consistent grading practices for students, teachers and all the other stakeholders. Looking back to the entire journey, it was an opportunity to know myself

better, and a chance to enhance personal competencies. Either one could not be achieved if I did not have the guidance and support from professors, staff, colleagues, and friends I met in the programme.

Firstly, I want to thank Professor Cheng, who gave me the chance to continue my academic study at Queen's and introduced to me the world of *grading*. Secondly, I want to thank my supervisor Professor DeLuca, who was always patient and provided professional guidance and instructions throughout my graduate study. Thirdly, I would like to thank Professor Klinger, who guided me in my first poster presentation; Professor Shulha, who taught me programme evaluation and offered me my first field practice experience; Professor Lee, Professor Bolden, Professor Luce-Kapler, and Angela, who enlightened me when I was confused. Besides, I would like to express my appreciation to the AEG group and EGSS community at Faculty of Education. It is an unforgettable experience to work with all the excellent and passionate colleagues! Especially, I want to thank Jia, Yi, Yongfei, Xiaoqian, Ulemu, Jess, Allison for offering me academic support and always being so nice to explain anything that I may not know well; to Esther, Stefanie, Jessica, Lorraine, Chi, Stefan, and Ian for helping me develop the research idea and improve my writing skills; to Meaghan, Jill, Maria, Judy, Luna, Jennifer, and Rebecca for just being such a sweet friend. Also, I would like to thank Marlene, Celina, Natalie, Brenda, Corinne, Peter, Amy, Maggie, Vicky and all the other staff and librarians in the faculty. Without your help and support, I cannot achieve the goal. Thank you!

Last but not the least, I want to thank Tibor, Victoria, Jackie, Daren, Kyra, Gordon, Fang, Elizabeth, Adeela, Estee, Manju, Kiran, Brian, Dr. Qin, Dr. Zhang, Ms.

Xu, Ms. Pan, Ms. Tan, Jingying, Chenning, Jiayan, Pingping, Yiran, Wanwan, Yifei, Jie, Chen, Moying, Si, Ting, Fang, Qing, Xiaochun, and Miaoqing. Thank you for your endless support and encouragement. And I would like to express my sincerest appreciation to my parents. Thank you for believing in me and thank you for being there all the time!

“Sapientia et Doctrina Stabilitas”—Wisdom and knowledge shall be the stability of thy times. At the end of one journey and the beginning of a new one, hope the Queen’s spirit, its motto and the memory of my MEd study will always motivate me to find the new “balance point” in every journey in the future. No matter how time flies.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	vii
List of Abbreviations	viii
Chapter 1 : Introduction	1
Chapter 2 : Literature Review	6
Chapter 3 : Methodology	20
Chapter 4 : Results	33
Chapter 5 : Discussion	50
References	58
Appendix: List of Policies	64

List of Tables

Table 1: Definitions for grades

Table 2: Selected provinces and cities

Table 3: Inductive analysis of policy documents (i.e., categorical congruence)

Table 4: Theme frequency for categorical concurrence dimension

Table 5: Grading suggestions for different subjects

Table 6: Alignment dimension: Range of knowledge

Table 7: Range of knowledge: Theme frequencies

Appendix A1: National policies

Appendix A2: Provincial policies

Appendix A3: School policies

List of Abbreviations

1. GOC: The Government of China
2. MOE: Ministry of Education
3. NPCSC: Standing Committee of the National People's Congress
4. DEB: District Education Bureau
5. EB: Education Bureau
6. ESC: Education Supervisory Committee
7. HB PES: Hangzhou Baochu Pagoda Experimental School
8. HCEMS: Hangzhou Caihe Experimental Middle School
9. HJMS: Hangzhou Jingfang Middle School
10. HWMS: Hangzhou Wenlan Middle School
11. HXPS: Hangzhou Xuejun Primary School
12. NEC: National Education Council
13. POE: Provincial Office of Education
14. PG: Provincial Government
15. RCPS: Rainbow City Primary School
16. SLPS: Shaoxing Luxun Primary School
17. SWMS: Shaoxing Wenlan Middle School
18. WBS: Wahaha Bilingual School

Chapter 1: Introduction

Grading, also known as marking, is a decision-making process used to summate student achievement with a mark or symbol (Marzano, 2000; Tomlinson, 2005; Brookhart, 2013; McMillan, 2001). Grading can be used to provide targeted and specific information, such as to grade individual assessments or to provide broad information based on a culmination of student achievement to inform achievement, selection, promotion, and placement decisions (Brookhart, 2004, 2013). In relation to classroom teaching and learning, a *grade* is generally understood as “a piece of information that attempts to report something about a student’s education” (Quinn, 2013, p. 5) that then combine “into one summary mark or symbol for a report period” (Brookhart, 2013, p. 257). Hence, the initial and primary purpose of grading is to communicate student achievement of learning (Brookhart, 1993, 2004; Marzano, 2000; Guskey, 2000, 2006; Lambating & Allen, 2002; Randall & Engelhand, 2010). However, this simple-looking metric for reporting student achievement is imbued with values, meanings, and multiple interpretations related to curriculum, teaching, learning, assessment, and a gamut of other potential factors that impact the grading process and resulting grade.

Contemporary research on grading presents the consistent finding that “academic performance is clearly the most important factor in grading students” (McMillan et al., 2002, p. 211; McMillan, 2001; McMillan, Myran, & Workman, 2002; Randall & Engelhard, 2008, 2009, 2010; Sun & Cheng, 2013). In addition to achievement, teachers may consider a mixture of non-achievement factors (e.g., effort, participation, improvement, and credit work) in the assignment of grades, especially with borderline

cases (McMillan, 2001; McMillan et. al., 2002; Randall & Engelhard, 2010). The inclusion of these diverse set of factors is highly variable across teachers (McMillan, 2001; Guskey, 2006; Cheng & Wang, 2007; Tierney, Simon, & Charland, 2011) Therefore, there is a possibility that only the teachers who assign the grades can fully understand their meaning. Other stakeholders' understanding will depend on how well the grade scales, rubrics, grading principles, and procedures are articulated during the reporting period and on the report card (Brookhart, 2013).

Furthermore, researchers have recognized that grading in schools is not a simple, technical decision-making process; instead, like all the other kinds of assessment, it can be deeply social and personal (Johnston, Guice, Baker, Malone, & Michelson, 1995; Black & William, 1998). Grading can be further challenged by the diversities of schools, subjects, classes, students and teachers (McMillan et al., 2002; Brown, 2003; Duncan & Noonan, 2007; Cox, 2011). The interpretation of a grade by various stakeholders is further based on a variety of factors: personal beliefs, backgrounds, professional knowledge, previous experiences, assumptions and attitudes, purposes and intentions, as well as the social, cultural and educational contexts (Waltman & Frisbie, 1994; Thomas & Oldfather, 1997; Ross & Kostuch, 2011; Sun, 2011; Brookhart, 2013).

Brookhart (2004), one of the leading researchers on grading, commented that the best grading systems should be designed for its communicative function to provide accurate, valid, and useful information on students' learning as well as clear statements on the factors that contribute to grade decisions.

Nevertheless, former researchers have recognized that potential misinterpretations and misalignments exist in the coupled processes of grade assignment and grade

interpretation. Without additional or further explanatory and descriptive statements about grades as well as grading process, other stakeholders (i.e., students, parents, administrative officers) may not clearly understand teachers' assessment of student achievement (Ross & Kostuch, 2011). Moreover, insufficient and ambiguous statements as well as inconsistent definitions of terms can cause misunderstanding, even confusions among stakeholders (Friedman & Frisbie, 1995; Munk & Bursuck, 2001). These layers of interpretation and variations in grades can affect individuals' judgments, ultimately causing validity concerns. As individuals make inferences from assigned grades as well as related determinants, which will affect grades' implication for action, and influence students' self and professional development in the future (Messick, 1989; Brookhart, 1991; Friedman & Frisbie, 1995; McMillan & Workman, 1998; Allen, 2005; Guskey, 2006; Sun & Cheng, 2014).

Although it is acknowledged that there is no single or so-called "right way" to grade (O'Connor, 2002), there are definitely some principles or guidance for assigning grades, which can make each grade "better serve the purpose of providing interested individuals with useful information" (Tomlinson, 2005, p.263), and minimize the possibilities of misinterpretation and invalidity. Educational systems develop and implement grading policies to provide teachers with principles and guidelines for making grading decisions. Grading policy documents may consider both pre-established academic factors (Brookhart, 2011) and other related factors, and provide teachers with criteria descriptions linked to corresponding learning qualities (Swan, Guskey, & Jung, 2014).

However, despite the intention to provide guidance to teachers, grading policies may provide mixed messages (i.e., diverse definitions, processes, and orientations to grading) as they are often issued at multiple levels within the educational system. Currently, the consistency and alignment of these policies across educational levels are largely unknown. But studies have shown that a lack of alignment may lead to further diversification in teachers' grading practices and greater confusions rather than consistent grading practices that yield dependable interpretations. For instance, when policies and supporting documents become hard to differentiate, teachers may simply ignore these documents (Webb, 1997) instead of trying to understand and remember the content or aligning interpretations with actual classroom practices (Perna & Davis, 2007; Squires, 2005, 2009). Evidence also suggests that when teachers do not understand a new policy, some teachers may assume that all related policies and standards align with each other (Squire, 2005 & 2009; Zheng, 2013). Others tend to rely on their previous experiences and personal understanding (Brookhart, 1993, 1994; McMillan, 2001; McMillan et al., 2002; Zheng & Ye, 2012). These reactions might lead other stakeholders (e.g., students, parents, other teachers, administrative officers) to doubt or question the credibility of the resulting grades (Tuten, 2007; Ross & Kostuch, 2011). Therefore, there is a need to investigate the alignment of grading policies across an entire educational system, to ensure that teachers are being provided with consistent messages about how and what to grade.

The purpose of this study is to examine the alignment of grading policies at varying levels in the Chinese elementary and secondary education system (i.e., national,

provincial, and school levels). Specifically, the following two questions will guide this study:

1. How do national, provincial, and school level policies describe grading?
2. What is the alignment between grading policy documents across two dimensions (i.e., categorical congruence and range of knowledge) at various educational levels (i.e., national, provincial, and school levels)?

An alignment methodology is used to respond to the two questions. Results from this study may contribute to the knowledge of grades and grading in the Chinese context, and may also serve as a reference for future grading policy and grading related studies.

Structure of the Study

This study is structured in five chapters. Chapter one presents a brief introduction of study background, the rationale, and major research questions. Chapter two presents a literature review that analyzes related empirical studies and provides a theoretical basis for this study. Chapter three outlines study's methodology (i.e., data sources and alignment method) that used to answer the research questions. Chapter four presents results related to the alignment of grading policies across system levels. The last chapter discusses key findings, study limitations, and provides recommendations for future research.

Chapter 2: Literature Review

In this chapter, previous empirical studies and published articles related to grades and grading practices are reviewed in order to provide the basis and background information for this study. The chapter begins by exploring the definition, meaning, and construction of grades. Then, studies focused on grading policies are reviewed. The chapter concludes with a section on grading in the Chinese education system.

Definition of *Grade* and *Grading*

The first use of *grade*, as a noun, can be dated back to the 1510s and its meaning was “degree of measurement” at that time. Later in 1807, *grade* had been given a more specific meaning related to measurement, which was “class of things having the same quality or value”. The meaning that became more closely to assessment started from the usage of “*numerical grades*”, and when letter scale had been introduced to grading practices, the word was given another meaning in 1886, which specifically represented the “letter-mark indicating assessment of a student’s work”. However, it was not until 1650 that the word *grade* had been used as a verb for the first time, and the meaning “assign a letter mark to” was not attached to *grade* until 1931 (Online Etymology Dictionary).

Contemporarily, the meaning of *grade* has become richer after a long time of development and definitions related to assessment have become more specific. Findings from online Oxford Dictionary are as follows:

Noun

1. A particular level of rank, quality, proficiency, intensity, or value;

1.2 chiefly *North American* A mark indicating the quality of a student's work

1.3 *British* An examination, especially in music

Verb

1.1 chiefly *North American* Give a mark to (a student or a piece of work)

(Online Oxford Dictionary)

From the above definitions, a general and literal meaning of *grade* is giving a mark, which can represent the quality or the level of students' academic performance. It shares the same idea that researchers once defined the word *grade* that was a mark or symbol representing students' achievements (Randall & Engelhard, 2009a; Brookhart, 2013), which clearly describe the aim and the function of this symbol. *Grading*, instead, emphasises the action and process when the word is used as a verb.

In the context of classroom assessment, *grading* (sometimes called *marking*) is one of the most common means that teachers use in their day-to-day teaching. It is a process that sums up student achievement with marks or symbols (Brookhart, 2013). Teachers use the phrase of "*grading papers*" when they scoring or rating individual assignments; they use "*grade(s)*" when referring to the final mark on the report card (Brookhart, 2004). Yet, neither *grade* nor *grading* is a simple symbol or "process of summing up student achievement with marks or symbols" (Brookhart, 2013, p.257). It is "a value-laden decision-making process" that particularly influenced by "values, and beliefs about teaching and learning" (Sun, 2011, p.44). Take the example Quinn (2013) used in her book *On Grades and Grading* (Table 1). When providing individuals with an uncompleted definition sentence, i.e., "A grade is a/an _____ of _____ relative to _____.", responses can be classified into five major categories based on the type of

information, i.e., quantification, symbol, description, evaluation and ranking. Besides, the finding also revealed that information type can affect the focus of grading process as well as the context that individuals and stakeholders use for evidence collection.

Table 1.
Definitions for grades

Type of info	Topic of info	Context for info
<i>A grade is a/an ...</i>	<i>Of ...</i>	<i>Relative to ...</i>
Quantification	A student's learning	A standard
Symbol	A student's skill level at certain time	A student's peers within a class, school, grade, or age level in a particular region
Description	A student's average skill level over a period of time	A student's starting point
Evaluation	A student's performance on an individual assessment task	All other possible outcomes
Ranking	A student's performance on a number of assessment tasks	A student's progress over a certain period of time

- *Note: From On Grades and Grading Table 1.1 (Quinn, 2013, p. 6)*

When defining *grade* and *grading* in an educational assessment context, the first priority should be providing clear and complete definitions, which does not only explain what a *grade* is but also articulate primary purposes and intentions for grading. Moreover, scales for *grade* and *grading* should be provided as well. Because ambiguous definitions will affect the construction of *grades*, which may lead to different grading focuses, supporting evidence, and eventually, it may result in inconsistent and blurry interpretations, and affect the use as well as consequences of *grades*. Thus, one of the

focuses in this study is finding out descriptions of *grade* and *grading* in policy documents in order to see how they have been defined in the Chinese context.

The Construction of Grades

The major debate in grading construction is whether to consider achievement factors only or not when assigning grades. Assessment researchers and experts suggest that in order to avoid misinterpretations and misunderstandings, *grades* should only reflect students' academic performance and achievement (McMillan et al., 2002; Randall & Engelhard, 2009b). Therefore, only achievement related factors should be considered. However, the reality is teachers initially want to assign the grades based on students' achievement factors only, or at least they agree with the idea that grades should only show students' academic achievement, they consider non-achievement factors in their actual practices due to different reasons (e.g., personal beliefs or preferences, experiences, grading policies, reporting forms, use and consequences, etc.; Randall & Engelhard, 2010; Sun, 2011; Brookhart, 2013).

In McMillan's (2001) study about secondary teachers' classroom assessment and grading practices, he surveyed among 1483 secondary teachers in Virginia, the USA. In the survey, McMillan (2001) asked question about factors and types of assessments that all the participant teachers used to determinate grades. The result showed that although academic achievement was consistently the most important component in teachers' grading practices, most secondary teachers used a multitude of factors in grading. The four major factors were academic achievement, academic enablers (e.g., effort, ability, improvement, and participation), external benchmarks, and extra credit and borderline cases.

Later, McMillan and his colleagues (2002) conducted a similar survey in Virginian elementary schools with Grade 3 and Grade 5 teachers. The results showed similar findings but revealed a great variance within schools that individual teachers' preferences were more important than differences between schools when determining grading practices (McMillan et al., 2002). This indicated the significant influence of teachers' understanding and interpretations about *grades* and *grading*. The reason behind this phenomenon can be attributed to different personal beliefs, knowledge, educational philosophies, experiences, attitudes, as well as purposes and use of the *grade* (Sun, 2011).

Except the individual differences among teachers and variances among schools, the reporting tools and process may also affect the *grade* construction as well. Friedman and Frisbie (1995) did a study to find out the relationship between the report card characteristics and the validity of grading information. They noticed that report cards from kindergarten to high schools seldom included philosophy and purpose statements; achievement symbols were attached with ambiguous descriptions, especially in kindergarten and elementary schools; the way to report non-achievement factors (i.e., effort, attitude, and conduct) varied above the elementary school level; in addition, the frequent use of computerized report card format also limited teachers' ability to describe students' performances adequately. In summary, setting and format of the report card may not allow teachers to leave comments and deliver messages the way they want (Allen, 2005; Tuten, 2007), which can affect the way how teachers select determinate factors before assigning the grade. These discrepancies and variances will ultimately affect the meaning of a grade.

The Meaning of Grades

Researchers have noticed from years of study that stakeholders understand and interpret *grades* and *grading* differently.

For student teachers, although they obtained basic knowledge of classroom assessment from professional trainings in school, and were aware the main purposes was “evaluating motivated students, communicating with parents, classifying students, and assessing their own effectiveness” (Brookhart, 2004, p. 434), their understanding about *grade* and *grading* was still limited to marking papers (Brookhart, 2004). For administrative officers, *grades* can represent the performance of students, the quality of teachers’ work as well as the performance of school programmes at certain period of time (Randall & Engelhard, 2010; Brookhart, 2011b; Liu, 2013).

For parents, *grades* are the evidence for children’s rewards, punishments or permission for extra-curricular activities (Randall & Engelhard, 2010), also means the classroom placement of their children (Brown, 2004). In 1994, Waltman and Frisbie did a survey comparing teachers’ and parents’ understanding on the same grade in Iowa, the USA. 16 Grade 4 mathematics teachers and 285 students’ parents participated in the study. The result found that although both math teachers and parents were not able to tell the differences between norm-referenced and criterion-referenced grading practices, parents normally believe the assigned grades were intend to include both achievement and non-achievement factors. Besides, parents’ also took that for granted that the average grade for every assessment was always lower than what teachers reported, That was, for example, when teachers reported an average grade of B, parents usually thought the average was C+ (Waltman & Frisbie, 1994). This finding revealed the concern that

parents had about grade inflation, which can affect the validity of teachers' grading practices.

Teachers' perceptions about the meaning of *grades* and *grading* were another research focus in many empirical studies. Among all the research, Brookhart (1993) and Sun and Cheng (2014) all did studies about the meaning that teachers assigned to a grade, using Messick's theory of validity as the framework.

In Brookhart's (1993) study, she recruited pre-service teacher candidates from the MEd programme in an American university to participate a scenario-based survey. Multiple choice questions about different grading scenarios and an open-ended question were provided to the participants. They chose the answers and explained the reason for the decisions they made in their own words in the open-ended question. After analysing all the responses, Brookhart (1993) found that the meaning of grade construction, for teachers, was closely related to the idea of student work: "Teachers used phrases like 'the work (s)he did' to describe grades or grading. *Performance* or *perform* were words often used, in the sense of something done or accomplished. To teachers, grades were something students earn; they were compensation for a certain amount of work done at a certain level" (p. 139). The result indicated that the factor of achievement was still an obvious major part for grading. But teachers had concerns about the fairness in grading process and thought really seriously about this issue. Brookhart (1993) thought it was because previous studies had found that teachers had dual roles of advocate and judge when assess students' performance. So, they had the concern about the consequences of the grades to the student, because the message delivered by the grade could have more interpretive and influential impact for students' future study.

Sun and Cheng (2014) carried out their study in a different context. The study had been conducted in 20 secondary schools in five school districts of a northern Chinese city. They surveyed 350 secondary school English language teachers about the meaning they associate with the assigned grades and the value judgments they make in grading. From their findings, the meaning of grades as well as the construct of grades among Chinese teachers was closely related to two concepts. The first one was the “judgment of students’ work in terms of effort, fulfilment of requirement, and quality” (p. 14). The second one was the “judgment of students’ learning in terms of academic enablers, improvement, learning process, as well as achievement” (p.14). These findings were consistent with previous studies that teachers considered various factors when assigning grades but “achievement is part of the construct, but not the whole of it” (p.14). For Chinese teachers, they particularly gave more weight to effort among all the non-achievement factors. Researchers attributed the reason to a common belief that most Chinese teachers had, which was learning depends upon effort rather than ability.

In order to minimize the discrepancies in *grade* interpretation, experts and authorities usually provide suggested recommendations to guide teachers’ classroom practices. However, fewer teachers would follow the instructions. The next section will discuss reasons that researchers find so far.

Recommendations and Teachers’ Classroom Practice

Discrepancy between teachers’ actual grading practices in the classroom and suggestions or advice recommended by experts is one of the issues noted by researchers and educators in previous literatures (Stiggins, Frisblie, & Griswold, 1989; Brookhart,

1993, 1994, 2004; Randall & Engelhard, 2010). Studies have been trying to explain the reasons behind this phenomenon.

Stiggins and his colleagues (1989) had attributed possible reasons to three aspects: a) recommendations may be still ideas and theories, which were lack of sufficient guidance, b) lack of consideration of actual classroom settings; c) lack of proper training to help teachers to master the new methods. Allen (2005) also stated that “inadequate education in valid assessment and grading principles and practices” (p.218) was the reason that affected the validity of most teachers’ grading practices in the classroom. Brookhart (1991), however, proposed another theory from teachers’ perspective. She pointed out the key reason was because teachers consider the uses and consequences of grades in their grading practices because they knew “grades are not only interpreted, they are used” (p. 35). Quinn’s (2013) findings were consistent with this idea as well. Table 1 shows that the word *grade* can have distinctive definitions when aiming differently, it also varies when understandings and interpretations have been given by different stakeholders.

Therefore, guaranteeing the alignment among different documents and support resources can help teachers conduct better and more efficient grading practice in the classroom, and improve the validity of assigned grades as well.

Grading in China

In order to improve the overall education qualities, narrowing the gaps between China and other developed countries, the past decades witness the major and radical reform among different levels of the Chinese education system (Peng et al., 2006; Ryan, 2011). According to previous literature, the Chinese elementary education alone has

experienced seven major curriculum reforms since the founding of the People's Republic of China in 1949 (Zhang, 2014). At the beginning of the twenty-first century, Chinese government launched its eighth curriculum reform and it is still under the implementation process (Liu & Kang, 2011). Among all the contents of each education reform, assessment and evaluation are always one of the major concerns as they are intimately related to the education quality and often be used as a direct referential factor when talking about the students achievements, as well as be widely used to facilitate education system under the present background of global changes and economic growth (Peng et al., 2006).

In a chronological review of the eighth curriculum reform, Liu and Kang (2011), as two key initiators and organisers of China's curriculum reform programme, summarized the major events and significant achievements throughout the five main stages of the reform. In addition, they provided reflections and evaluations to each of the very aspect. Regarding the specific content of assessment and evaluation at the elementary education level, the authors provided a logical line started from the establishment of reform guideline at the first planning and developing stage (1996—1998) to the stage of re-reflection, re-interpretation and further implementation since 2007 (Liu & Kang, 2011). From their description, assessment had being included into the aims of the reform, together with the other five areas (i.e., curriculum goals, structures, content, implementation, and administration), to the blueprint document after a long-time discussion and debates at the first stage. At that time, “assessment” was still a concern mainly at the administrative level, focusing on the educational outcomes and qualities that the reform could bring to the society. Yet, with the deepening of the practice, the

reality as well as the emerging issues and obstacles in the process made all the researchers and educators in the reform programme to think a much deeper and more fundamental question instead of simply focusing changing the curriculum system physically as previous reforms did, i.e., transforming from “mainly including changes to curriculum subjects, the increase or decrease of class times, the updating of teaching content, adjustments to teaching requirements and the updating of textbooks” (p.33). Many now consider the actual impact of the reform, and think about questions such as the following: What are the expectations for students’ first six to nine, or even 12-year experiences of schooling? What will be the mission for contemporary Chinese basic education? Within their statements, there were three times when the topic of assessment and evaluation raise up the debate and discussion both in the team and in public. The first time was at stage three (2001—2004) after the secondary school curriculum programme had been finalized and the pilot programmes had been implemented in the compulsory education curriculum. At this time, the reform team faced a challenge of ensuring the development was proceeding as expected as they realized that an effective appraisal and monitoring system was needed but the structure of it was still under exploring. The second time occurred when criticism rose from both education sector and the academic arena to question the quality of education under the new curriculum standard. Their concerns, which were surprisingly consistent, focused on the students’ academic results: “Teachers and parents have consciously or unconsciously compared the achievements of students under the new curriculum with the precious one in terms of academic marks”, which formed a general public believe that “students’ foundational knowledge and skills are not as solid as previously”, and “the quality of education has dropped” (p.30). This

lasting issue directly resulted in the reflection on the evaluation of contemporary basic education in China at stage five (since 2007). From Liu and Kang's (2011) perspective, they thought the reason why reform confronted oppositions could be categorized into two major factors. The first one was the lack of a more "democratic, open, scientific, equal, dialogic and consultative" (p.39) dialogue mechanism between education and academic filed. The second one was the efficiency of policy and reform implementation, which required consistent and continuous support and contribution from various shareholders.

The Development of Grading in Chinese Elementary Education

Since the first imperial examinations took place in the Han Dynasty, China started its history on testing and examination, and the impact of the examination system has greatly influenced the Chinese society ever since (Cheng & Curtis, 2010). Same as the examination system, neither the influence of the grades and grading system to individuals' learning experiences can be ignored.

During the time of ancient imperial examination, most of the examinations and tests were taken in a written form. All the papers written by examinees was graded by their writing style and content, those who succeeded in the exam were given certain names based on their ranking and would be offered opportunities to go to more intensive examinations at the higher level (Cheng & Curtis, 2010). Since 1860s, the reform of Chinese modern education sought for new testing and examination formats at the higher education level, and at the same time, assessment and evaluation system at the elementary education level was experiencing reforming and changing as well.

According to Yang's (2009) review of the Chinese examination system, the development of grades and grading system at the elementary education level had gone

through several stages. The first stage started after the enactment of “Provisional Regulations on Primary Schools” on March 18, 1952. In this stage, students’ academic and non-academic achievements in a term had been graded separately, and students’ final grades included three sections, the score of each subject, the score for physical education, and the score for daily behaviours or conduct. The calculation guidelines in the regulation indicated that for a single subject score in one term, 60% of the total score derived from students’ quizzes, tests and mid-term examinations. The other 40% came from the final examination at the end of the term. The subject score for one academic year would be the mean of scores in two terms. For the grading scale, teachers in most schools were required to use percentage grading and the benchmark will be 60. Only a few of the authorized schools approved by the department of education could use the five-level grading; score of behaviours or conduct was based on students’ daily performance, teachers used five-level scale and written feedbacks in grading (Provisional Regulations on Primary Schools, 1952).

The second stage happened during the time of “Cultural Revolution” (1966-1976) when examinations had been suspended in primary and secondary schools. The scores for students were mainly from “democratic assessment within in the school” (p.241), which can be really subjective if no valid and reliable criteria were provided.

Before the development of grades and grading system moved into third stage, the issue of students’ heavy academic load in primary school was widely discussed. “*Agreement of Instruction Programme in Full-time Six-year-type Primary Schools*” and “*Certain Regulations on Reducing Learning Load for Pupils in Primary Schools*” (p.242) were published in 1984 and 1988, respectively. The two documents caused a reform on

examination setting, times and subjects in primary school. Moreover, the new principle “schools should not publicize students’ scores and make rankings of students” (p.243) had been documented as a written rule, which result in five-level scale (i.e., excellent, very good, good, satisfactory, and ordinary) and written feedback for students’ academic achievement in their final grades. Since then, the grade and written feedback method, or “rank plus remarks” (p.247) referred to in Yang’s (2009) review, has been adopted widely by nearly all the primary schools in China.

Chapter 3: Methodology

The purpose of this study is to understand the alignment of grading policies across levels of the Chinese education system (i.e., national, provincial and school). Drawn from previous research, selected documents were analysed deductively in relation to two alignment dimensions: (a) categorical concurrence, and (b) range of knowledge. In this chapter, the methodology used in this study is described within following three sections: (a) alignment methodology, (b) alignment model used in this study, and (c) policy sources.

Alignment Methodology

Alignment methodology had been used in previous educational research to examine the consistency of educational policies, curriculum expectations, and assessment structures (Webb, 1997; Martone & Sireci, 2009; DeLuca & Bellara, 2013). In using an alignment methodology to examine curriculum expectations in relation to assessment tasks, Webb (1997) noted, “aligning assessment with expectations can improve the efficiency and effectiveness of the education system” (p. 2). Further, alignment studies can help set priorities and allocate limited resources more effectively; map students’ learning progress throughout the teaching and learning process; reduce unnecessary repetitive assessments; guarantee the effective planning of the system (Webb, 1997). However, several challenges have been identified when enacting an alignment methodology. These challenges include the growing number of policy and curriculum documents, repetitive expressions for expectations and assessments in various documents, inconsistency of using common terms or languages, and constant changes in the policy environment (Squire, 2005, 2009; Webb, 1997).

Webb (1997), a prominent alignment researcher, published a monograph outlining criteria for alignment studies related to curriculum expectation and assessment alignment in the subjects of mathematics and science. In this monograph, he explained the importance of alignment and defined it as “the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward student learning what they are expected to know and do” (p.4). He summarized three common approaches for aligning expectations and assessments. The first approach involved sequential development, which could help “to assure vertical alignment between the expectations and public opinion” (p.8). The disadvantages of this approach were time-consuming, and the “the process of developing expectations and assessments is not linear or sequential” (p.9).

The second approach involved an expert review. This approach had a formal review process conducted by a number of specialists to determine the degree to which curriculum expectations were mapped onto assessment tasks. Central to this approach was ensuring that expert reviewers had “clear direction of what their task is and what it is not... any content analysis or analysis of agreement among complex elements can be very detailed” (p.10).

The third approach included detailed document analyses, which was used to judge alignment through inductive and deductive coding and analyses. Practical issues that needed to be considered during this approach involved document selection, category selections, and impacts or biases from coders’ previous experiences. Category selections would influence analysis while impacts and biases would affect the final results. More specifically, how coders defined each category and how they selected criterion

represented their understandings, which could be varied. Thus, a common metric was needed when using document analysis (Webb, 1997).

To guide the document approach, Webb (1997) initially introduced five major criteria, i.e., content, students, instruction, and application, developed from earlier standards review and alignment studies. However, later in 2007, Webb proposed a new method for conducting document alignment studies. He established dimensional categories for aligning policies, curriculum expectations, and/or assessments. These categories included: (a) categorical concurrence, (b) depth of knowledge consistency, (c) range of knowledge correspondence, and (d) balance of representation. A full alignment between policies, curriculum expectations, and/or assessments happened only when all four criteria met. More specifically, the full alignment only happened when “sufficient number of items allocated to each standard with an appropriate level of complexity and coverage and without overemphasizing any one content area” (p.16). At the same time, five practical issues, including acceptable level of item numbers, items distribution related to a standard, range of knowledge in standard content coverage, balance of representatives among different standards, and the dynamic depth-of-knowledge level across grades, still need to be considered to ensure the quality and validity of results from an alignment study.

In order to evaluate the alignment degree between assessments and content standards in the American context, Bhola and his colleagues (2003) analyzed several alignment methods to determine three groups of alignment studies based on their level of complexities. These groups were: low, moderate, and high complexity models. Bhola et al. (2003) pointed out those models in the low complexity group “form the basis of

studies to provide content evidence to enable appropriate interpretation and use of test scores” (p.22). The moderate complexity group (e.g., Survey of Enacted Curriculum [SEC], Council for Basic Education [CBE], etc.), however, was created under the circumstance that content standards and assessment items were required for a dual perspectives exam in both content match and cognitive complexity match. The high complexity group included models developed by Webb (1997), La Marca et al. (2000), Achieve (2001), and Porter (2002). Bhola (2003) et al. discussed each model and concluded that “the more complex the model used to align assessments to standards, the less likely that items that match the standards will be found” (p.24). No matter which kind of models was used, the research results always depend critically upon the definition of the criteria, and the degree of alignment was not always correlated with the number of analysis dimensions.

Bhola et al. (2003) also identified several challenges across alignment methods. Consistent with Webb’s finding in 1997, Bhola and his colleagues identified the central challenge of using alignment methods was criteria specificity. They noticed that the variation of documents’ structures and requirements could affect the interpretation of certain components, i.e., some document components were designed to be only interpreted holistically or specifically, but stakeholders may not realize it. Furthermore, the narrowly stated expectations in the documents only provided minimum number of assessment items or tasks, which may not match the amount of specificity those policy makers expect. Thus, identified items may not cover the multidimensional curriculum expectations and policy statements, which could influence the actual assessment process and the result of expectations. Last but not the least, the efficiency of training provided

for coders and experts before they began the alignment task was a challenge as well. Training was essential and researchers needed to provide high-quality training as well as be aware of the overly generous tendency when participants looking for the matched items.

Martone and Sireci (2009) also did a comparative study to examine Webb's (1997), Achieve's (2001) and Poter and Smithson's (2002) alignment models. Different from Bhola (2003) et al., Martone and Sireci not only described differences and similarities of the three models, but also differentiated the three models by their application. They proposed, "Webb approach is useful to understand a degree of alignment" (p.1356) and provided "helpful information to determine what next step was needed in the process of revising the assessments and /or the standards" (p.1356). The SEC model, developed by Poter and Smithson (2002), was useful for enacting curriculum studies as the model "allows for alignment analyses across textbooks, professional development tools, and many other aspects of the educational process" (p.1356). Achieve's (2001) model, although it was not clearly delineate, could be used to study broader assessment qualities as related to curriculum standards. When choosing among these three models for research purposes, Martone and Sireci (2009) note that researchers need to depend on their study purposes and consider the availability of resources (e.g., financial, time, personnel resources, etc.). As Martone and Sireci (2009) regarded the alignment research as an ongoing process, they strongly recommended researchers and educators carefully consider the information types they need before selecting a proper method. Moreover, they also suggested that in order to provide information on future alignment process and acceptable dimensions, more studies on how the results of

alignment research are used (i.e., views from participants as well as interpretations from stakeholders) are required.

DeLuca and Bellara (2013) conducted an alignment study to examine educational policies, standards for assessment, and teacher education curriculum in order to reflect on the current state of assessment education in the United States. Their study aimed to “explore the alignment of the *intended* curricula with standards and policies for preservice assessment education” (p.360). An analytic alignment approach developed from previous alignment models and combined with a moderate complexity approach was used. DeLuca and Bellara (2013) re-defined the three dimensions (i.e., categorical concurrence, depth of knowledge, and range of knowledge) from Webb’s (1997) model and combined it with two subject matter experts (SMEs) to code related data. By analysing teacher education policy, standards for teachers practice in the area of student assessment and evaluation, and assessment education course syllabi from Florida, they found that the alignment of categorical concurrence and depth of knowledge were greater between standards documents and course syllabi than between policy documents and course syllabi. However, the representations of certain themes (e.g., assessment purposes, assessment for learning, communication of assessment results, classroom environment and assessment, etc.) were different across the documents. Further, the degree of emphasis may vary in different levels of documents.

Given the prominence of alignment designs across educational research, the following section articulates the specific alignment method used in this study.

Alignment Model

To understand the degree of alignment across grading policies within three levels of the Chinese education system, a revised alignment model was used in this study. The model was a moderate complexity model (Bhola et al., 2003) and drawn from Webb's document analysis approach (1997, 1999) involving two alignment dimensions: (a) categorical concurrence, and (b) range of knowledge.

The dimension of categorical concurrence typically aims to identify content consistency across different documents (Webb, 1997). In this study, policy documents were first uploaded to the Nvivo 10 software. Then, contents of grading and assessment policies were inductively analyzed by topic headings, subheadings, definitions of each topic, and explicit articulations of grading policy statements. Content statements within the policy documents were highlighted and categorized by identified codes, whose numbers were automatically calculated by the Nvivo 10 software and saved for frequency calculations. For example, one national policy stated that students' academic performance should be graded in primary school (*Advancing Assessment and Testing Reform in Primary And Secondary Education*, 2002), and another school policy also mentioned that grade and grading practices were recommended for teachers' classroom assessment (*Teaching Routine for Shaoxing Luxun Primary School*, 2003). These two statements would be coded the same, and the number of frequency would be automatically marked as 2 under this code. In the end, all the codes were compared and 53 codes were initially identified. These codes were thematically grouped into 8 content themes related to either *assessment* or *grading* (Table 3). The theme frequency was then calculated by different

educational levels (i.e., national, provincial and school) and the result was presented in Table 4.

The original intention for the other dimension, *range of knowledge*, was to analyze policy alignment from the perspective of its knowledge span (i.e., the specificity of the statements to inform grading practice). The requirement of a complete alignment was that all ranges of expected knowledge needed to be covered within each specified documents (Webb, 1997). The dimension, range of knowledge, in this study continued to explore the specificity of each topics and categories (i.e., to analyse how specific the grading related topics and categories were in each policy). The level of this dimension were drawn from the three levels used in DeLuca & Bellara's (2013) research (i.e., instructional, educational and global). The three levels represented the specificity of grading guidance and suggestions that teachers were informed by policy documents. Definitions for each level were presented in Table 6. All the identified codes in the dimension of categorical congruence were re-coded in order to identify instructional, educational, and global contents. The numbers of codes within each theme were automatically calculated by the Nvivo 10 software, and a matrix table (Table 7) was made in order to show the theme frequency in relation to the three alignment levels (i.e., instructional, educational and global) at different educational levels (i.e., national, provincial and school).

All inductive and deductive coding was conducted by the principal researcher. A subset of data was independently coded by a second rater to ensure analysis reliability. A high degree of inter-rater reliability was observed.

Policy Sources

Policies used in this study were selected from regions in China that participated in a trial implementation on education quality comprehensive evaluation reform in primary and secondary schools. In a statement published by the General Office of Ministry of Education (MOE) in 2013, thirty regions have been identified as within the experimental areas for primary and secondary schools' education quality comprehensive evaluation reform. Excluding three municipalities (i.e., Beijing, Shanghai, and Chongqing) and three autonomous regions (i.e., the Guangxi Zhuang Autonomous Region, the Ningxia Hui Autonomous Region, and the Xinjiang Uygur Autonomous Region), fifteen provinces and twenty-three cities were selected as the initial subject pool for this study. Then, after a further search for the accessible documents in publically available literature, fourteen provinces and four cities were selected within the provincial level. At the school level, however, due to the limitation of time and available online documents, only nine schools from Zhejiang province were selected for this study. Details of the provinces, cities and schools are presented in Table 2.

Table 2.
Selected provinces and cities

Educational Levels		Information			
Province		Guangdong	Guizhou	Heibei	Henan
		Hubei	Hunan	Jiangsu	Jiangxi
		Liaoning	Shandong	Shanxi	Sichuan
		Yunnan	Zhejiang		
City		Shijiazhuang	Changsha	Nanchang	Hangzhou
District		Jiangan District, Hangzhou, Zhejiang			
		Shangcheng District, Hangzhou, Zhejiang			
Schools	Primary Schools	Hangzhou Xuejun Primary School			
		Rainbow City Primary School			
		Shaoxing Luxun Primary School			
	Middle Schools	Hangzhou Jingfang Middle School			
	Hangzhou Wenlan Middle School				
	Shaoxing Wenlan Middle School				
	Experimental School	Hangzhou Baochu Pagoda Experimental School			
		Hangzhou Caihe Experimental Middle School			
		Wahaha Bilingual School			

In this study, all the publically accessible policy documents were collected directly from the following four sources:

1. Major search engines (e.g., Google, Google Scholar, Baidu, etc.). Using keywords (e.g., grading, grades, *dengji* [等级], *dafen* [打分]), title keywords and full title search (e.g., grading policy, grading policies in China) in both Chinese and English to identify related documents.
2. Published articles and public databases (e.g., CNKI, Wanfang Data). Using keywords and subject search to identify published articles on grades, grading, grading policies, assessment and evaluation in the Chinese context and then looking for policy documents referred in those articles.

3. Government websites (e.g., Ministry of Education, Provincial Department of Education). Searching for authorized documents that have been given public access.
4. School official websites. Looking for published documents at the school level.

Two kinds of policy documents were used in this study: (a) grading and assessment related policies at different educational level (e.g., national, provincial, school, etc.) in China, and (b) curriculum standards for four major subjects (i.e., Chinese Language, Math, English, and Science) at compulsory education stage (i.e., Grade 1 to Grade 9) in China in which grading practices were discussed. Details for the two document categories are as follows:

Policy documents.

Policy documents were sorted by educational levels. At the national level, eleven major documents were identified, including three development plans (i.e., *The Twelfth Five-Year Plan for National Economic and Social Development of the People's Republic of China*, *The Twelfth Five-Year Plan for China's Educational Development*, and *The Mid- and Long-Term Plan for Education Reform and Development of the People's Republic of China*), two laws (i.e., the *Teachers Law of the People's Republic of China* and the *Compulsory Education Law of the People's Republic of China*), two professional standards for primary and secondary teachers (MOE, 2011), one document for basic education curriculum reform (MOE, 2001), one statement of *Advancing Assessment and Testing Reform in Primary and Secondary Education* (MOE, 2002), one statement and a framework of *Advancing Reform of Primary and Secondary Education Quality*

Evaluation (MOE, 2013). The three development plans and two laws set up foundations for educational evaluation and assessment practices in the Chinese contexts as well as provide guidance for other policy making. The remaining documents indicated trends and focus on current evaluation and assessment issues, which impact teachers' classroom grading practices.

At the provincial level, three different kinds of documents were collected and analysed from selected areas. These policy documents included the *Provincial Twelfth Five-year Plan for Education*, the *Provincial Mid- and Long-term Plan for Education Reform and Development*, and the *Implementation Plan for Education Quality Comprehensive Evaluation Reform in Primary and Secondary Schools*. Policy documents from the four selected cities had been grouped into the provincial level for analysis. At the school level, school policies, development statements, and other grading and evaluation related documents (e.g., teaching routine, teaching regulations, school regulations) were identified. Documents from two school districts were also categorized into the school level for analysis. Details of all policy documents used in this study can be found in Appendix A.

Curriculum standards.

In China, Chinese Language, Math, English, and Science are four major subjects in primary and secondary schools. Curriculum standards of the four subjects usually provide direct principles and guidelines for teachers to prepare and carry out teaching activities, including grading. In this study, the newest curriculum standards authorized by the Chinese Ministry of Education (MOE) were used in these four subject areas. Standards were downloaded from the MOE website directly. In total, five curriculum

standards documents were downloaded because the curriculum standard for Science had two documents based on grade level (i.e., one is for Grade 3-6 and the other one is for Grade 7-9). Detailed information of all the curriculum standards can be found in Appendix.

Chapter 4: Results

In this chapter, results of the analysis of Chinese grading policies are presented in the following order. First, codes and themes inductively identified from all documents are presented with a description of each theme and its associated codes. These themes describe the primary content within each policy document and are used in the categorical congruence and range of knowledge alignment dimensions. Theme frequencies related to categorical concurrence and range of knowledge dimensions are then presented. The range of knowledge table, together with the theme frequency table for the categorical concurrence dimension, are used to respond to the two research questions that drive this study and indicate the degree of alignment among grading policies, standards and other grading documents at different educational levels (i.e., national, provincial, and school levels).

Inductive Thematic Analysis: Theme and Code Descriptions

To determine potential codes within the categorical concurrence dimension, each policy, standard, and document used in this study was inductively coded for its content (Patton, 2002). Throughout the coding process, two major categories were identified: *assessment* and *grading*.

This study noticed that in all of the collected documents, grading was highly entwined with the concept of assessment in the Chinese context, with a 0.65 co-occurrence rate. For example, one national document, *The Announcement of Promoting Assessment and Testing System Reform in Primary and Secondary Schools* (2002) stated that assessment should focus on both students' improvements and assessment results; it should benefit both students and teachers' personal development; and in primary schools,

students' academic performances were suggested to be graded. At the provincial level, a government policy titled "Teaching Advice for Teachers in Primary and Secondary Schools in Zhejiang Province" stated, "teachers are encouraged to use assessment methods integratively: combining formative and summative assessments; using interview and paper-and-pencil tests together; giving encouraging comments with grades" (p.7). These statements are consistent with findings in previous studies as grading is considered as a subset of assessment (McMillan & Workman, 1998; Brookhart, 2004, 2008; Duncan & Noonan, 2007; Sun, 2010). Since no official policies or documents were found specifically relating to the issue of grading in China, a study on grading practices needs to embrace broader understandings about assessment within the Chinese educational context. Therefore, in this results chapter, I examine policies related to *grading* and *assessment*, establishing distinctions between the two concepts where appropriate.

After deductively coding all data, 53 codes were initially identified. These codes were then thematically grouped into 8 content themes related to either *assessment (general)* or *grading*. For example, the codes *obey supporting standards/policy documents, maintain balanced education development, sense of diversity, objectivity/efficiency/transparency, validity, fairness, providing specific suggestions for future teaching and learning* were all grouped into the theme of *assessment principles*. *Grading is one of the rights for teachers, using grading for students' academic performance, the purpose of grading, using grades as the major means to report academic performance in elementary schools, applying grading/grades in secondary and even higher education* were grouped into the theme of *Grading Principles*. The eight identified themes ranged from understanding fundamental principles and purposes to

communicating grading. Details on the identified themes and their associated codes can be found in Table 3 below. Descriptions of these themes suggest multiple dimensions to grading purposes and processes based on documents from across educational levels. In the following section of results, I examine the alignment of these content themes across document types (i.e., alignment of categorical congruence).

Table 3.
Inductive analysis of policy documents (i.e., categorical congruence)

Theme	Associated codes	Description of theme
1. Assessment Principles	<ul style="list-style-type: none"> • Obey supporting standards/policy documents • Maintain balanced education development • Sense of diversity (students, methods, goals, etc.) • Objectivity/Efficiency/Transparency (tools, design, process, etc.) • Validity (information and result) • Fairness • Providing specific suggestions for future teaching and learning 	Standards and policy documents that need to be followed. Fundamental and basic principles that need to be considered when conducting assessments
2. Assessment Purposes	<ul style="list-style-type: none"> • Encourage students and explore their potentiality • Improving students' academic performance • Providing feedback/helping students' further development • Enhance academic knowledge • Planning/adjusting teaching plans • Monitoring teaching process • Diagnosis/examination/selection 	Involves general assessment purposes for different stakeholders
3. Assessment Contents/Dimensions	<ul style="list-style-type: none"> • Subject knowledge and thinking patterns • Problem solving and other practical skills • Learning process and method/participation • Emotion/attitude/value • Academic achievement/outcome • Self-cognition • Innovation knowledge and thinking • Communication and cooperation 	Dimensions and perspectives that teachers need to consider when assessing students' performance

Table 3 continued

4. Communicate and Use Assessment Results	<ul style="list-style-type: none">• Using proper forms for results presentation (scores, grades, descriptions, etc.)• Using results efficiently for timely feedback/communication/analysis• Do not use the result for any ranking purpose Do not using the result as the only criterion for teacher and school evaluation• Recommend grading/matching standards for summative assessment	Involve recommendations for assessment result formats, interpreting, communicating and using the results.
5. Grading Principles	<ul style="list-style-type: none">• Grading is one of the rights for teachers• Using grading for students' academic performance• The purpose of grading• Using grades as the major means to report academic performance in elementary schools• Applying grading/grades in secondary and even higher education	Basic and general interpretations about grading, its purpose and usage. In the meantime, indicating the undergoing reform of grading practices.
6. Grading Contents/Dimensions	<ul style="list-style-type: none">• Knowledge and ability/study method• Practice and implementation• Emotion and attitude/habit/personal effort• Daily/mid-term/final academic performance (attached with different percentage)• Academic progress• Group work/cooperation• Suggest rubrics/criterion/examples for different subjects/skills	Dimensions and perspectives that teachers need to consider when grading and assigning grades to students.
7. Grading Practice	<ul style="list-style-type: none">• Keep daily/phased/progress records for students' academic performance• Sense to the variety of stakeholders and their positive involvement• Aim to arouse and maintain students' learning enthusiasm• Objective/reliable/humaneness• Grade students' actual academic performance• Undergoing reform at the elementary education level (e.g., grading scales, reporting format etc.)	Suggestions for teachers to consider when assigning grades, including considering students' daily performances in the classroom; the impacts of other stakeholders;

Table 3 continued

8.Grading Result Communication	<ul style="list-style-type: none"> • The scale of letters/proficiency categories are recommended • Aim to encourage students and communicate their performance objectively • Provide timely communication with other stakeholders (students/parents) • Provide timely feedback for teaching and learning • Do not using the result for any ranking/selection purpose • Do not using the result as the only criterion for teacher and school evaluation • Report card forms are recommended 	Providing recommended scales and reporting formats for teachers' grading practices while articulating the results using and communication among stakeholders. Besides, re-emphasizing the purpose of grading.
--------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Categorical Congruence

The first research question of this study aims to find out how grading is described in published documents at different educational levels (i.e., national, provincial, and school). Table 4 presents the relative theme frequencies for each level of documents based on the content themes derived from the inductive analysis.

Table 4
Theme Frequency for Categorical Concurrence Dimension

Theme	Educational Levels		
	National	Provincial*	School*
1 : Assessment Principles	0.13	0.08	0.02
2 : Assessment Purposes	0.40	0.10	0.03
3 : Assessment Contents/Dimensions	0.25	0.20	0.03
4 : Communicate and Use Assessment Results	0.13	0.45	0.03
5 : Grading Principles	0.01	0.01	0.02
6 : Grading Contents/Dimensions	0.05	0.00	0.49
7 : Grading Practice	0.00	0.02	0.16
8 : Grading Result Communication	0.03	0.15	0.22

Based on an analysis of categorical congruence, several alignment trends are evident. At the national level, emphasis across policy documents is placed on the broad

topic of *assessment*. More specifically, these documents focus on *assessment purposes* (0.40) and *assessment contents/dimensions* (0.25). This finding indicates that policy makers aim to provide educators with a clear explanation of the fundamental purposes of assessment whilst providing general guidance for selecting appropriate assessment processes for teachers' classroom contexts. For example, two announcements about primary education reform and assessment reform in primary and secondary schools stated, the educational function of assessment should play an effective role, "(assessments) should be used to discover and develop students' potentials, understand their needs, and help them enhance self-awareness thereby build self-confidence" (*Basic Education Curriculum Reform*, 2001). The new assessment system should aim at students' development and include clear and precise assessment contents, criteria, methods and improvement suggestions (*The Announcement of Promoting Assessment and Testing System Reform in Primary and Secondary Schools*, 2002). The other category, *grading*, and its related themes, only reflected a small proportion of data within national policies, with a focus on three grading themes mainly (i.e., *grading principles*, 0.01; *grading contents/dimensions*, 0.05; and *grading result communication*, 0.03). Among all the selected documents, it was common to find grading statements such as:

- "Grades are recommended for reporting students' academic performances, especially for primary school students" (*Curriculum Standards for English*, 2011; *The Announcement of Promoting Assessment and Testing System Reform in Primary and Secondary Schools*, 2002).
- "Grading results are private and cannot be used for ranking and selection purposes" and "it cannot be used as the major criterion for teachers' or schools' performance

evaluation” (*The Twelfth Five-Year Plan for China’s Educational Development*, 2012).

The analysis of provincial policies showed a similar emphasis on assessment themes but with an overall lower frequency (0.83). Provincial level government policies developed a major focus on the theme of *communicate and use assessment results* (0.45), which was nearly quadruple the amount at the national level. This suggests that, on the one hand, provincial policies pay more attention to communication and the effective use of assessment results in order to achieve goals and purposes of assessments. For instance, one provincial policy stated, “the assessment results should not be announced to public”, it should be used to help teachers “identify and analyse major problems and issues in students’ study” and “looking for improving strategies” (*Teaching and Learning Principles for Primary and Secondary Schools in Changsha*, 2015). On the other hand, this finding also reflects the profound influence of assessment in the Chinese context. Since the beginning of implementing testing and examination as the major assessment methods in the Chinese history, the results were always attached with ranking, selection and decision making purposes, which would eventually affects one person, even one family’s future development (Cheng, 2010). Therefore, clarifying the usage and communication of the results may help teachers, students, parents, administrative officers and other related stakeholders have a new perspective of assessment and its results whilst interpreting and using them more properly and effectively. Moreover, it is consistent with the undergoing reform trend mentioned in the national document that “the educational function of assessment should be used more effectively” (*Basic Education Curriculum Reform*, 2001), and this goal can be fulfilled if stakeholders use results more

meaningfully. For example, one document from Shandong Province pointed out explicitly that assessment results should be used rationally. The results should be used to “gather the information about students’ current performance in the study, as well as help teachers adjust pedagogical methods and improve teaching and learning qualities” (*Principle of Teaching for Primary and Secondary Schools in Shandong Province*, 2015).

The data analysis process also revealed two other important findings related to the broad theme of assessment. The first finding focused on the use of assessment results. This study found that provincial policies in China were consistent with those at the national level, emphasising that assessment results, especially those that represent students’ academic performance, cannot be used for ranking, selection or other evaluation, rewarding or punishment purposes (*Regulations for Compulsory Education in Guizhou Province*, 2012; *The Mid- and Long-Term Plan for Education Reform and Development Of Henan Province*, 2013; *The 12th 5-Year Plan for Shandong’s Educational Development*, 2014; *Regulations of Compulsory Education in Zhejiang Province*, 2010). The second finding pertains to the stakeholders’ rights. Analysis showed Guangdong was the only province identified that encouraged junior high school students to save their most satisfied academic performance grades from repeated junior high school academic performance assessment for future academic use (e.g., applying senior high schools, making admission decisions); it aimed to fulfil different students’ potential and developmental needs (*The 12th 5-Year Plan for Guangdong’s Educational Development*, 2011). In addition, the Guangdong provincial government permitted “teachers, students and parents have the right to submit written appeal if they are not satisfied with the results. Schools should accept and respond without delay” (*Criterion for Standardized*

Schools at Compulsory Education Level in Guangdong Province, 2013). These two findings re-emphasize the importance of communicating and using assessment results properly and effectively.

In relation to grading, provincial policies covered three of the four themes: *grading principles (0.01)*, *grading practice (0.02)*, and *grading result communication (0.15)* but with little emphasis. Interestingly, in comparison to national level policies, contents within the theme of *grading practice* started to be covered at the provincial level. Though the frequency was low, this finding suggests that these policies are beginning to give teachers guidance on suggested grading protocols. For example, the *Teaching Advice for Teachers in Primary and Secondary Schools in Zhejiang Province* (2009) had presented suggestions for different subjects. Table 5 presents direct quotes from provincial curricular documents.

Table 5
Grading Suggestions for Different Subjects

Subjects	Primary Education Level	Secondary Education Level
1. Chinese Language	<ul style="list-style-type: none"> • Consistent with the teaching goal • Focusing on students' actual literacy competencies and communicating skills • Signing grades with encouraging comments 	<ul style="list-style-type: none"> • Focusing on content and key points of each unit; reflecting students' actual literacy competencies and communicating skills
2. Math	N/B	<ul style="list-style-type: none"> • By the end of each unit, get to know each student's proficiency level, accept students' individual differences • For slow learners, teachers should lower the criterion and offering re-assessment/grading when needed in order to protect students' self-confidence • Phased assessment/grading at the end of the term needs to reflect every student's improvement/achievement

Table 5 continued

3. English	<ul style="list-style-type: none">• The primary goal is arousing students' learning enthusiasm and initiatives• Based primarily on students' interests, attitude and communication ability in daily English learning in the classroom• Evidence can be collected from different aspects (e.g., study portfolio; survey; interview; self/peer evaluation; parents feedbacks/comments)• Grade 3 and Grade 4: grading on their abilities using English to complete tasks; Grade 5 and Grade 6: grading on their actual listening, reading skills, the ability to apply what they learn into practice, and the proper writing habit	<ul style="list-style-type: none">• Involving different stakeholders (e.g., teachers, students, peers, parents) in the process• Teachers are suggested to assign grades based on different aspects (e.g., performance in class; portfolios; survey, interview)
4. Science	<ul style="list-style-type: none">• Grading are recommended for assessing students' scientific knowledge• By the end of each unit, phased records are recommended; by the end of each term, summative assessments are recommended	<ul style="list-style-type: none">• Try to provide an overall view of students' scientific proficiencies

In contrast to national and provincial policies, the topic of assessment was not the primary emphasis at the school level as observed by its relatively small proportion in the document contents (overall frequency: 0.11). Analysis of school level policies yielded a notable shift towards grading. The primary emphasis of school level documents was the theme of *grading contents/dimensions* (0.49). Hence school level policies provided teachers with practical guidance on grading criterion and dimension setting. Specifically,

there were two types of grading policies that have been identified. The first one was associated with tests and examinations, especially associated with final examinations. For example, one school policy stated, “for Math, grading should cover oral calculation, calculation, basic mathematic knowledge, application and operation. And for Chinese Language, it should include (1) *pingying*, characters, words, sentences; (2) listening, speaking, reading skills, and (3) the proficiency of literacy. All the results will be presented in grades” (*Teaching Principles for Luxun Primary School*, 2003). The other type represented more comprehensive grades that students had by the end of each term. For instance, one school policy stated, “the grade for one academic year is based on students’ grades from both examination and non-examination subjects. Furthermore, students’ daily performance in school will also be counted into the final grades” (*Students Status Management Policy for Caihe Experimental School*, 2012).

The themes of *grading practice* (0.16) and *grading result communication* (0.22) were a secondary emphasis at the school level, and the frequencies were higher than both national and provincial levels. These findings suggest school level policies aim to present more detailed information about grading practice and provide direct instructions for teachers. For example, some of the policies stated, “students’ academic performance should include grades of both examination and non-examination subjects” (*Students Status Management Policy for Caihe Experimental School*, 2012), “four-level scale (i.e., excellent, good, pass, and fail) is recommended for examination and daily assignment assessment, comments can be added when needed” (*Teaching Principles for Subjects in Compulsory Education in Shangcheng District*, 2011). Other policies also mentioned that “students’ performance and grades should be saved with a digital record” and focus on

their learning process (*Working Plan for Dean’s Office in 2014-2015 Academic Year*, 2014). In all, “grading should be fair, objective and with no biases” (*Teaching Principles for Luxun Primary School*, 2003), and it “should be encouraging as well” (*Teaching Principles for Subjects in Compulsory Education in Shangcheng District*, 2011). More importantly, it should be used to support students’ development and teachers’ pedagogical adjustment (*Teaching Routine for Luxun Primary School*, 2003; *Teaching Advice for Primary and Secondary Schools Teachers in Zhejiang Province*, 2009; *Students Status Management Policy for Caihe Experimental School*, 2012).

Range of Knowledge

All the selected official policy documents were analysed for their range of knowledge to determine the degree of content specificity by educational levels. Analysis of these documents provides in-depth understanding on how specific the statements are for the adoption and implementation of grading and assessment policies at different educational levels. In this study, there are three levels of content specificity (i.e., instructional, educational, and global). In general, these levels represent whether teachers get broad, general and more educational information about grading, or have more practical, applicable and instructional guidance from official policy documents. The definitions of the three levels are presented in Table 6.

Table 6
Alignment Dimension: Range of Knowledge

Range of knowledge level	Coding criteria
1. Instructional level	Detailed instructions for teachers daily grading practice in classrooms (e.g., recommended criterions for different subjects, rubrics for overall grades/one specific content; scales)

Table 6 continued

2. Educational level	Statements contain descriptions of grading that aim to educate teachers as well as to prepare them with basic knowledge and skills before grading students' performance in the classroom (e.g., options of grading scales and reporting formats; general aspects that can be considered when assigning a grade)
3. Global level	General and broad statements about grading/assessment, the purposes, the undergoing reform and interpretations of grading under the overall concept of assessment (e.g., it is teachers' right to grade students' academic performance; grading aims to encourage students)

Table 7 presents results based on the range of knowledge analysis for policy documents at national, provincial, and school levels. Findings from this analysis is consistent with results observed related to categorical congruence: Comparatively, national and provincial documents developed greater emphasis on global and educational levels related to general *assessment* topics while provincial and school policies express content at educational and instructional levels related to *grading*.

Table 7.
Range of Knowledge: Theme Frequencies

Categorical Congruence	Educational Level	Instructional	Educational	Global
1. Assessment principles	National	0.00	0.15	0.50
	Provincial*	0.00	0.09	0.55
	School*	0.00	0.33	0.67
2. Assessment purposes	National	0.00	0.05	0.57
	Provincial*	0.00	0.21	0.64
	School*	0.00	0.00	0.83
3. Assessment contents/dimensions	National	0.00	0.45	0.08
	Provincial*	0.00	0.57	0.07
	School*	0.00	0.20	0.00
4. Communicate and use assessment results	National	0.00	0.26	0.32
	Provincial*	0.00	0.50	0.17
	School*	0.00	1.00	0.00

Table 7 continued

5. Grading in general	National	0.00	0.00	1.00
	Provincial*	0.00	1.00	0.00
	School*	0.00	0.25	0.75
6. Grading contents/dimensions	National	0.13	0.00	0.00
	Provincial*	0.00	0.00	0.00
	School*	0.28	0.47	0.00
7. Grading practice	National	0.00	0.00	0.00
	Provincial*	0.00	1.33	0.00
	School*	1.11	0.63	0.04
8. Grading result communication	National	0.00	0.50	0.00
	Provincial*	0.00	0.86	0.05
	School*	0.32	0.39	0.13

At the global level, there were significant areas of alignment across themes of: *assessment principles* (national: 0.50; provincial: 0.55; school: 0.67), *assessment purposes* (national: 0.57; provincial: 0.64; school: 0.83), and *grading contents/dimensions* (national: 0.00; provincial: 0.00; school: 0.00). The themes of *assessment contents/dimensions* and *communicate and use assessment result* were only represented at national and provincial levels while *grading results communication* was only identified at provincial and school levels with a minimal and low frequency respectively. The theme of *grading principles* was evident at both national and school levels with a high frequency (i.e. 1.00 at national level; 0.75 at school level) but not evident globally at a provincial level. Additionally, *grading practice* was only marginally mentioned in school documents at a global level (0.04). Overall, these findings suggest that, globally, policies and documents from three educational levels contain more broad and conceptual information within the theme of *assessment*. It is common to find

statements related to principles and purposes, which provide baselines as well as background knowledge for teachers' practice in classroom.

At an educational level, results indicate areas of strong alignment among national, provincial, and school policies for the themes of *assessment principles*, *assessment contents/dimensions*, *communicate and use assessment results*, and *grading result communication*. Across each of these themes, the analysis indicates that policies at national, provincial, and school levels have a fairly consistent frequency at an educational level for the themes of: *assessment principles* (national: 0.15; provincial: 0.09; school: 0.33), *assessment contents/dimensions* (national: 0.45; provincial:0.57; school: 0.20), *communicate and use assessment result* (national: 0.26; provincial: 0.50; school: 1.00), and *grading result communication* (national: 0.50; provincial: 0.86; school: 0.39). Additionally, the results also reveal a same changing pattern of the frequency in both *assessment contents/dimensions* and *grading result communication*, which is the content frequency increases from national to provincial level, however, drops from provincial to school level. This overall trend reflects that provincial documents focus more on providing referenced assessment content and suggestions for communicating grading results.

At the instructional level, results show a strong alignment within the topic of *assessment* and within the theme of *grading in general* (national: 0.00; provincial: 0.00; school: 0.00). For the rest of the themes, results in Table 7 show that *grading contents/dimensions* has been slightly covered at both national level (0.13) and school level (0.28). The content of *grading result communication* minimally covered in school level documents (0.32). However, *grading practice* has been addressed with a high

frequency (1.11) at school level, which consistent with the finding in categorical concurrence dimension that school level policies contains more practical and applicable information.

Overall, the table of range of knowledge indicates that alignment exists across most of the content themes. More specifically, alignment can be found in the general topic of *assessment* at both instructional and educational level. Misalignments were also observed across the range of knowledge dimensions. For instance, the themes of *assessment purposes*, *grading principles*, *grading contents/dimensions* and *grading practices*, all have very different frequencies at the educational level. Furthermore, comparing to the content coverage at both national and provincial level in the themes of *assessment contents/dimensions*, and *communicate and use assessment result*, the theme of *grading result communication* was only represented at provincial and school levels at the global level.

Apart from the misalignment, the results indicate that grading related contents are mostly covered at provincial and school levels. Provincial policy documents also contain a fair amount of assessment related contents with national policy documents developed the most coverage of assessment contents. In relation to the theme of *grading practice*, the analysis reveals that this is the only theme with misalignment at all three range of knowledge levels (i.e., instructional, educational and global). Results show that national policies do not have information on grading practices. However, in provincial and school policies, grading practices have greater coverage at educational levels. Additionally, school level documents have been identified with a minimal frequency (0.04) at global level but with a significant frequency (1.11) at instructional level.

Summary

In conclusion, findings from this alignment analysis show that national policy documents provide general statements to teachers about *assessment*, which normally set the conceptual boundaries for assessment related activities, as well as providing fundamental principles for policy making at lower levels. These general statements on assessment also provide a global basis for grading policies at lower system levels. While provincial authorities have the autonomy to decide their own assessment and grading policies, the majority of provincial documents examined in this study aligned closely with national level policies related to assessment. It is evident that for specific guidance on grading, teachers in China should turn to local level policies, which have the greatest number of statements at the most instructional level for grading purposes and processes.

Chapter 5: Discussion

Grades are symbols that contain information about student achievement based on specified learning period (Brookhart, 2013; Quinn, 2013). They can be represented by scales of letters, numbers, points, or written comments. However, assigning a grade is a “value-laden decision-making process” (Sun, 2010, p.2) involving multiple factors (i.e., a mixture of achievement and non-achievement factors). In order to maintain the focus on academic achievements, grading policies are developed to link explicit criteria and procedures for teachers to follow. Further, grading policies contain definitions and orientations towards grading. Yet, multiple levels of the educational system can contain multiple documents with multiple definitions and orientations, which are especially true in the case of Chinese education. Thus, the lack of alignment among policy documents can result in expanding discrepancies and diversifications in teachers’ grading practices. In addition, a lack of central guidance about grading may lead to a complex and diverse record-keeping system that eventually cannot efficiently support students’ learning (Black & Wiliam, 2005), and may be challenged for the validity as well.

By examining the alignment of current grading policies in the Chinese education system, this study examined (a) the description of grading across three different educational levels (i.e., national, provincial, and school), and (b) the degree of alignment across categorical congruence and range of knowledge dimensions at these levels based on the specificity of identified codes and themes. Based on results from this study, several conclusions can be drawn related to definitions of grades/grading across policy documents, purpose and content of grades, and the degree of policy alignment in China.

Definitions of *Grade* and *Grading*

Findings showed that policy documents focused more on assessment in general in the Chinese context. Descriptions for *grade* and *grading* were normally drawn on the concept of assessment, which was consistent with early studies (McMillan & Workman, 1998; Brookhart, 2004, 2008; Duncan & Noonan, 2007; Sun, 2010). The intentions and purposes for both assessment and grading had been well recognized, which could be attributed to a) the historic impact of examination and testing in China (Cheng & Curtis, 2011), and b) the ongoing curriculum and assessment reform in China (Zheng, 2005; Zhang, 2014).

According to the findings, there were two general descriptions about *grade* and *grading* that could be identified in all the documents. The first description was using *grade* as a symbol to represent students' academic performance. Normally, policy documents recommended teachers to use letter scale (i.e., A, B, C, and D) or three-level scale (i.e., excellent, good, and pass) for grading practices, especially at the elementary education level. This description reflected the general understanding that existed across previous studies (Marzano, 2000; McMillan, 2001; Brookhart, 2013; Swan et al., 2014).

The second description was using *grade* as a summative symbol that combined both academic and non-academic factors for students' final grade, which was another focus that had been widely discussed in previous studies (McMillan et al., 2002; Tomlinson, 2005; Ross & Kostuch, 2011; Sun & Cheng, 2014). The differences of descriptions led to different purposes, contents and dimensions that teachers may consider in their practices. It reemphasised the importance and necessity of giving a clear and specific definition.

However, this study found that there was little definition about *grade/grading* represented in national and provincial policy documents. Rather, national level policies described loosely the purposes of grades and the responsibility for teachers to grade students' learning. For example, one of the policy documents in the national level stated that assessing students and assigning *grades* for students' academic work was one of teachers' rights and responsibilities (*Teachers Law*, 1994). Other descriptions of *grade* and *grading* appeared in all the related documents described them as a suggested method used for teachers' assessment practices. For instance, one national policy stated that primary school students should be graded for their academic performance (*Advancing Assessment and Testing Reform in Primary and Secondary Education*, 2002). School level documents offered more articulated descriptions of grades/grading and their purposes. One school level document indicated that the assessment system in school aimed to provide a comprehensive assessment for students' all-round development, and their academic performance would be graded (*School Regulations for Hangzhou Jiangfang Middle School*, 2010). Another school level policy was given more specific guidance on teachers' grading practices. It stated that grades should be used for grading students' daily assignments and final examinations (*Teaching Regulations for Shaoxing Luxun Primary School [math]*, 2003). As evident from these statements, documents failed to provide clear and comprehensive definitions for *grades* and *grading*.

Purposes and Related Contents of *Grading*

Statements about purposes and content considerations for grading were evident across policy documents, mainly at national and school levels. One national statement indicated, assessment should be used to help students' and teachers' development, and

guaranteed the implementation of all-round education in the school system (*Advancing Assessment and Testing Reform in Primary and Secondary Education*, 2002). Another national document stated, assessment should not only focus on academic aspect but try to motivate students' potentials (*Basic Education Curriculum Reform [trial implementation]*, 2001). Grading, as one of the major means of assessment, should be used to help improve teaching method, examine educational outcomes, and ensure the teaching and learning qualities in school (*Teaching Routine for Shaoxing Luxun Primary School*, 2003).

About *grade* and *grading* related contents, national, provincial, and school policies all tried to cover contents from all the four themes (i.e., grading principles, grading contents/dimensions, grading practices, and grading result communication), which would provide a global understanding about grade and grading for different stakeholders. However, the focus among all the documents was still grading contents, and using grades for student achievement communication. This emphasis was consistent with Brookhart's (2004) idea about *grading*, which should be designed with a communicative function and system to provide accurate information about students' performance in school.

The Degree of Policy Alignment

Using the range of knowledge dimension, this study found that, overall, points of alignment and misalignment existed across policy documents from the three educational levels (i.e., national, provincial and school). A dominant trend in the alignment was national policies intend to give more general statements on grading while practical guidance was more commonly and easily found in the school level. Provincial policy

documents were more like the connection between the national and school level documents, and covered fairly equal amount of both general and practical contents.

The finding was consistent with the study published by DeLuca and Bellara (2013). Yet, in this study, there were potentially more profound reasons contributing to areas of alignment and misalignment. As Zheng (2013) stated in his study, due to the Chinese political system, the sensitive academic performance policy normally decided by the national administrative department (e.g., Ministry of Education), and policy documents published in lower educational level would try to be consistent with the national documents: “most of the time, provincial and local documents were just the explanation or complementation of the national documents” (p. 68). Therefore, the general consistency and alignment of the documents remained, but those documents may not match the variation among different regions and areas. And finally, schools and teachers may just rely on their understandings and experiences to assess and grade students in the classroom (Zheng & Ye, 2012).

Apart from the alignment, misalignment still existed in the content specificity dimension across the policy documents. For example, the curriculum of math presented that the criterion of assessing students’ calculation ability was the number of calculation one student could complete per minute (*Mathematics Curriculum Standards of Compulsory Education*, 2011), whereas the school level teaching routine just listed the major dimensions that teachers need to consider when assigning grades but without more detailed descriptions (e.g., criteria, rubrics). Continue using math as an example, one school policy stated that students’ performance would be graded by the end of each term. Final grades would include oral calculation, calculation, basic math knowledge and

application (*Teaching Regulations for Shaoxing Luxun Primary School* [math], 2003) but no detailed criteria had been presented in the document. Thus, further study may need to use a more complex alignment model to analyze the alignment and consistency across documents.

Limitations

Though this study created a general picture of grading in the Chinese context, there were some limitations. The major limitation for this study was related to the policy documents and the accessibility of school level documents and grading policies. As a MEd project, all the collected standards, policies and other related documents in this study were based on publically available sources. During the data collection process, it was noticed that though some of the documents used in the study were identified from previous empirical studies, not all the cited documents had been saved in paper version, or uploaded to related websites with a digital version. Furthermore, some of the departments and schools only had documents which were issued at an earlier time. For future studies, more recent documents, especially at the school level, should be collected by contacting or visiting officers and administrators in those departments and organizations.

Another limitation in this study was related to the regional representation of data. Due to the limited time and available policy documents, only fourteen provinces and nine schools from one province had been included at the provincial and school levels respectively. For subsequent studies, a larger data sample representing additional regions would be valuable.

Finally, the alignment framework used in the study only covered two major dimensions from Webb's (1997) original alignment framework. An alignment framework that also includes enacted practices would present an even more comprehensive account of grading policies within the Chinese context.

Future Research and Implications

Building upon this study, it would be valuable to research the following aspects of grading through subsequent studies.

The first aspect is the definition of *grade* and *grading*. Previous empirical studies had found that stakeholders hold different opinions about the purpose and function of grades and grading, which could affect their grading practices and the use of grading results. Policy documents were supposed to serve as the instructional and educational guidance for stakeholders, and used to minimize the communication gap among different stakeholders. However, according to findings from this study, published grading policy documents in China did not present clear definitions for grade and grading. Additionally, *grade* and *grading* related descriptions varied among different educational levels (i.e., national, provincial, and school). Thus, it will be helpful if future studies can be designed to find out how different stakeholders (e.g., policy makers, administrative officers, teachers, parents, and students) understand *grade* and *grading*, and how they interpret the grading policies in the Chinese contexts.

The second aspect is expanding the geographic scope of the study to include additional regions and data sources. Cities and provinces have discrepancies among each other, and involving more areas and schools in the Chinese context could provide better

understanding of grades and grading practices, as well as a more specific grading policy alignment situation in China.

Finally, the third aspect for future research is using a more complex alignment model to study the alignment of grading policy documents. As Bhola (2003) pointed out that the more complex the alignment model was the less likely that items would be found match the standards. Misalignment in content specificity had been identified from this study. A more complex model can be used to further explore the degree of alignment across policy documents in different educational levels with more specific dimensions, which can present a in depth description of the grading policy landscape in the Chinese education system.

References

- Allen, J. D. (2005). Grades as valid measure of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218–223.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: methods and issues. *Educational Measurement: Issues and practice*, 22(3), 21–29.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–144, 146–148.
- Brookhart, S. M. (2011). *Grading and learning: Practices that support student achievement*. Bloomington, IN: Solution Tree Press.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123–142.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279–301.
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, N.J: Pearson.
- Brookhart, S. M. (2011). The use of teacher judgment for summative assessment in the United States: Weighed in the balance and (often) found wanting. Presented at the Invited Research Seminar Teachers' Judgments within Systems of Summative Assessment: Strategies for Enhancing Consistency, Oxford, UK.
- Brookhart, S. M. (2013). Grading. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257–271). Thousand Oaks, CA: Sage.

- Brown, G. T. L. (2003). *Teachers' Conceptions of Assessment* (PhD Thesis). University of Auckland, NZ. Retrieved from <https://researchspace.auckland.ac.nz/bitstream/handle/2292/63/02whole.PDF?sequence=6>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318.
- Cheng, L., & Curtis, A. (Eds.). (2011). *English language assessment and the Chinese learner*. New York, NY: Routledge.
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4) 356–372.
- Duncan, C. R., & Noonan, B. (2007). Factors Affecting Teachers' Grading and Assessment Practices. *The Alberta Journal of Educational Research*, 53(1), 1–21.
- Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement*, 55(1), 5–26.
- Fung, Y. (1952). *A history of Chinese philosophy* (2nd ed.). Princeton, NJ: Princeton University Press.
- Grade. (n.d.-a). In *Online Etymology Dictionary*. Retrieved from http://www.etymonline.com/index.php?allowed_in_frame=0&search=grade&searchmode=none
- Grade. (n.d.-b). In *Online Oxford Dictionary*. Retrieved from http://www.oxforddictionaries.com/us/definition/american_english/grade
- Guskey, T. R. (2000). Grading policies that work against standards...and how to fix them. *NASSP Bulletin*, 84(620), 20–29.

- Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, 87(9), 670–675.
- Johnston, P., Guice, S., Baker, K., Malone, J., & Michelson, N. (1995). Assessment of teaching and learning in “literature-based” classroom. *Teaching and Teacher Education*, 11(4), 359–371.
- Lambating, J., & Allen, J. D. (2002). How the multiple functions of grades influence their validity and value as measures of academic achievement. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, AL, USA.
- Liu, J., & Kang, C. (2011). Reflection in action: Ongoing K-12 curriculum reform in China. In J. Ryan (Ed.), *Education reform in China: Changing concepts, contexts and practices* (Vol. 69, pp. 21–40). New York, NY: Routledge.
- Liu, X. (2013, May). *Investigating factors influencing grading decisions among teachers of Chinese to speakers of other languages* (MEd Thesis). Queen’s University, Kingston, Ontario. Retrieved from <http://qspace.library.queensu.ca/handle/1974/8050>
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332–1361.
- Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McMillan, J. H. (2001). Secondary teachers’ classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20, 20–32.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers’ classroom assessment and grading practices. *Journal of Educational Research*, 95(4), 203–213.
- McMillan, J. H., & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Retrieved from <http://files.eric.ed.gov/fulltext/ED453263.pdf>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.

- Munk, D. D., & Bursuck, W. D. (2001). What report card grades should and do communicate: Perceptions of parents of secondary students with and without disabilities. *Remedial and Special Education, 22*(5), 280–287.
- O'Connor, K. (2002). *How to grade for learning: linking grades to standards* (2nd ed.). Arlington Heights, IL: SkyLight Professional Development.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Peng, W. J., Thomas, S. M., Yang, X., & Li, J. (2006). Developing school evaluation methods to improve the quality of schooling in China: A pilot “value added” study. *Assessment in Education: Principles, Policy & Practice, 13*(2), 135–154.
- Perna, D. M., & Davis, J. R. (2007). *Aligning standards & curriculum for classroom success* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Quinn, T. (2013). *On grades and grading: Supporting student learning through a more transparent and purposeful use of grades*. Lanham, MD: Rowman & Littlefield Education.
- Randall, J., & Engelhard, G. (2009a). Differences between teachers' grading practices in elementary and middle schools. *The Journal of Educational Research, 102*(3), 175–185.
- Randall, J., & Engelhard, G. (2009b). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement, 46*(1), 1–18.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372–1380.
- Roach, A. T., Niebling, B. C., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessment: Implications and applications for research and practice. *Psychology in the Schools, 45*(2), 158–176.

- Ross, J. A., & Kostuch, L. (2011). Consistency of report card grades and external assessment in a Canadian province. *Educational Assessment, Evaluation and Accountability*, 23(2), 159–180.
- Ryan, J. (2011). *Education reform in China: Changing concepts, contexts and practices* (Vol. 69). New York, NY: Routledge.
- Squires, D. A. (2005). *Aligning and balancing the standards-based curriculum*. Thousand Oaks, CA: Corwin Press.
- Squires, D. A. (2009). *Curriculum alignment: Research-based strategies for increasing student achievement*. Thousand Oaks, CA: Corwin Press.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14.
- Sun, Y. (2011). Grading in language assessment: A literature review. In M. Ingersoll & R. L. Kapler (Eds.), *Graduate student symposium selected papers* (pp. 44–65). Kingston, ON: Queen's University.
- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343.
- Swan, G. M., Guskey, T. R., & Jung, L. A. (2014). Parents' and teachers' perceptions of standards-based and traditional report cards. *Educational Assessment, Evaluation and Accountability*, 26, 289–299.
- Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practice: "That's my grade. That's me." *Educational Psychologist*, 32(2), 107–123.
- Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum*, 75, 210–227.
- Tomlinson, C. A. (2005). Grading and differentiation: Paradox or good practice? *Theory Into Practice*, 44(3), 262–269.

- Tuten, J. (2007). "There's two sides to every story": How parents negotiate report card discourse. *Language Arts*, 84(4), 314–324.
- Waltman, K. K., & Frisbie, D. A. (1994). Parent's understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223–240.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Retrieved from <http://files.eric.ed.gov/fulltext/ED414305.pdf>
- Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states. Washington, DC: Council of Chief State School Officers. Retrieved from <http://eric.ed.gov/?id=ED440852>
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7–25.
- Yang, M. (2009). *Educational system in china*. Hangzhou, China : Zhejiang University Press.
- Zhang, Y. (2014). Reforms and review of China's basic education curriculum [中国基础教育课程改革现状与评述]. *Divineland* [神州], (15), 287.
- Zheng, D. (2005). Review of China's curriculum reform (since 1949) [新中国课程改革的历史回顾]. *Education and Vocation* [教育与职业], (13), 49–51.
- Zheng, D. (2013). Developing teachers' assessment literacy: From the perspective of academic assessment policies [发展教师评价素养：学业评价政策的视角]. *Shanghai Research on Education* [上海教育科研], (2), 67–70.
- Zheng, D., & Ye, S. (2012). The Survey on Classroom Assessment Knowledge of Primary and Secondary School Teachers and Its Origin: based on Zhejiang Province [中小学教师课堂评价知识及其来源的研究——基于浙江省的样本调查]. *Research in Educational Development* [教育发展研究], (20), 68–78.

Appendix: List of Policies

Appendix A1. *National Policies*

Administrative Level	Document ID	Document Title	Author/ Publisher	Year of Publication	Affiliated Region
National	N1	Teachers Law of the People’s Republic of China	NPCSC	1994	
	N2	Basic Education Curriculum Reform (trial implementation)	MOE	2001	
	N3	Advancing Assessment and Testing Reform in Primary and Secondary Education	MOE	2002	
	N4	Compulsory Education Law of the People’s Republic of China	MOE	2006	
	N5	The Mid- and Long-Term Plan for Education Reform and Development of the People’s Republic of China	MOE	2010	
	N6	Elementary School Teachers' Professional Standards (trial implementation)	MOE	2011	
	N7	Middle School Teachers' Professional Standards (trial implementation)	MOE	2011	
	N8	The Twelfth Five-Year Plan for National Economic and Social Development of the People's Republic of China	GOC	2011	China
	N9	The Twelfth Five-Year Plan for China's Educational Development	MOE	2012	
	N10	Advancing Reform of Primary and Secondary Education Quality Evaluation	MOE	2013	
	N11	The Framework for Primary and Secondary Schools’ Education Quality Evaluation (trial implementation)	MOE	2013	
National <i>(Curriculum Standards)</i>	CS1	Science Curriculum Standards of Grade 3—Grade 6	MOE	2007	
	CS2	Chinese Curriculum Standards of Compulsory Education	MOE	2011	
	CS3	Mathematics Curriculum Standards of Compulsory Education	MOE	2011	
	CS4	English Curriculum Standards of Compulsory Education	MOE	2011	
	CS5	Science Curriculum Standards of Junior High School Education	MOE	2011	

Appendix A2.
Provincial Policies

Administrative Level	Document ID	Document Title	Author/Publisher	Year of Publication	Affiliated Region
Provincial <i>(Including regional documents)</i>	P1	The Mid- and Long-Term Plan for Education Reform and Development of Guangdong Province	POE	2010	Guangdong Province
	P2	The Twelfth Five-Year Plan for Guangdong's Educational Development	PG	2011	
	P3	Further Promoting the Development of Balanced and Standardized Compulsory Education in Guangdong Province	POE	2013	
	P4	The Standard of Standardized Schools in Compulsory Education in Guangdong Province	POE	2013	
	P5	The Mid- and Long-Term Plan for Education Reform and Development of Guizhou Province	POE	2011	Guizhou Province
	P6	The Compulsory Education Regulations of Guizhou Province	POE	2012	
	P7	The Specialized Plan for Guizhou's Educational Development	POE	2015	
	P8	The Mid- and Long-Term Plan for Education Reform and Development of Hebei Province	POE	2010	Hebei Province
	P9	The Twelfth Five-Year Plan for Provincial Economic and Social Development of Hebei Province	PG	2011	
	P10	The Implementation Plan for Education Quality Comprehensive Evaluation Reform in Primary and Secondary Schools in Shijiazhuang	EB	2014	Shijiazhuang/Hebei Province
	P11	The Twelfth Five-Year Plan for Henan's Educational Development	POE	2012	Henan Province
	P12	The Mid- and Long-Term Plan for Education Reform and Development of Henan Province	POE	2013	
	P13	The Mid- and Long-Term Plan for Education Reform and Development of Hubei Province	POE	2011	Hubei Province
	P14	The Compulsory Education Regulations of Hubei Province	POE	2010	
	P15	The Plan for Strengthening Education in Hunan Province	POE	2011	Hunan Province
	P16	The Specialized Plan for Promoting the Development of Balanced Compulsory Education in Hunan Province	PG	2012	

Appendix A2 continued

Provincial <i>(Including regional documents)</i>	P17	The Twelfth Five-Year Plan for Hunan's Educational Development	POE	2012	Hunan Province
	P18	The Five-Year Plan for Monitoring and Evaluating the Fully Implementation of Education for All-round Development in Changsha	Changsha ESC	2015	Changsha/ Hunan Province
	P19	The Teaching Routine for Primary and Secondary Schools in Changsha	EB	2015	
	P20	The Mid- and Long-Term Plan for Education Reform and Development of Jiangsu Province	POE	2010	
	P21	The Education Development Plan for Jiangsu Province	PG	2011	Jiangsu Province
	P22	The Key Indicators for Developing High Quality and Balanced Compulsory Education	POE	2012	
	P23	The Mid- and Long-Term Plan for Education Reform and Development of Jiangxi Province	POE	2011	Jiangxi Province
	P24	The Twelfth Five-Year Plan for Jiangxi's Educational Development	POE	2012	
	P25	The Teaching and Management Regulation for Primary and Secondary Schools	EB	2014	Nanchang/ Jiangxi Province
	P26	The Mid- and Long-Term Plan for Education Reform and Development of Liaoning Province	PG	2010	Liaoning Province
	P27	The Twelfth Five-Year Plan for Liaoning's Educational Development	PG	2012	
	P28	The Compulsory Education Regulations of Shandong Province	POE	2009	
	P29	The Twelfth Five-Year Plan for Shandong's Educational Development	POE	2014	
	P30	The Guidance for Improving the Implementation of Academic Examination and Comprehensive Quality Evaluation System in Junior High Schools	POE	2015	Shandong Province
	P31	The Mid- and Long-Term Plan for Education Reform and Development of Shandong Province	POE	2015	
	P32	The Teaching Standards for Primary and Secondary Schools in Shandong Province	POE	2015	
	P33	The Mid- and Long-Term Plan for Education Reform and Development of Shanxi Province	POE	2010	
P34	The Twelfth Five-Year Plan for Shanxi's Educational Development	PG	2012	Shanxi Province	
P35	The Standards for Compulsory Education Schools in Shanxi Province	POE	2013		

Appendix A2 continued

	P36	The Education Development Plan for Sichuan Province	PG	2011	
	P37	Further Promoting the Development of Balanced Compulsory Education in Sichuan Province	PG	2013	Sichuan Province
	P38	The Mid- and Long-Term Plan for Education Reform and Development of Sichuan Province	POE	2013	
	P39	The Mid- and Long-Term Plan for Education Reform and Development of Yunnan Province	POE	2011	
	P40	The Twelfth Five-Year Plan for Yunnan's Educational Development	POE	2012	Yunnan Province
Provincial <i>(Including regional documents)</i>	P41	Further Promoting the Development of Balanced Compulsory Education in Yunnan Province	POE	2013	
	P42	The Guidance of Teaching and Management for Compulsory Education in Zhejiang Province	POE	2009	
	P43	The Teaching Suggestions for Primary and Secondary School Teachers	POE	2009	
	P44	The Compulsory Education Regulations of Zhejiang Province	POE	2010	Zhejiang Province
	P45	The Twelfth Five-Year Plan for Zhejiang's Educational Development	POE	2011	
	P46	The Mid- and Long-Term Plan for Education Reform and Development of Zhejiang Province	POE	2015	
	P47	The Mid- and Long-Term Plan for Education Reform and Development of Hangzhou	EB	2011	Hangzhou/ Zhejiang Province

Appendix A3.
School Policies

Administrative Level	Document ID	Document Title	Author/ Publisher	Year of Publication	Affiliated Region
School (Including district documents)	S1	The Suggestions for Academic Assessment in Jianggan District (for Math, English and Chinese)	DEB	2007	Jianggan District/ Hangzhou/ Zhejiang Province Shangcheng District/
	S2	The Regulations for Primary School Management	NEC	1996	Hangzhou/ Zhejiang Province Shangcheng District/
	S3	The Teaching Regulations for Primary and Secondary School Teachers in Shangcheng District	DEB	2011	Hangzhou/ Zhejiang Province SLPS/Shaoxing/ Zhejiang Province
	S4	Teaching Routine for Shaoxing Luxun Primary School	SLPS	2003	SLPS/Shaoxing/ Zhejiang Province
	S5	Teaching Regulations for Shaoxing Luxun Primary School (for Math, English and Chinese teachers)	SLPS	2003	SLPS/Shaoxing/ Zhejiang Province
	S6	Teaching Regulations for Shaoxing Wenlan Middle School	SWMS	2008	SWMS/Shaoxing/ Zhejiang Province
	S7	School Regulations for Hangzhou Jingfang Middle School	HJMS	2010	HJMS/Hangzhou/ Zhejiang Province
	S8	The Educational Reform and Ddevelopment Plan for 2011-2015	HBPEs	2011	HBPEs/Hangzhou/ Zhejiang Province
	S9	Teaching Regulations for Rainbow City Primary School	RCPS	2011	RCPS/Hangzhou/ Zhejiang Province
	S10	School Regulations for Hangzhou Caihe Experimental Middle School	HCEMS	2012	HCEMS/Hangzhou/ Zhejiang Province
	S11	The Teaching Regulations for Xuejun Primary School	HXPS	2013	HXPS/Hangzhou/ Zhejiang Province
	S12	Working Plan for Dean's Office in 2014-2015 Academic Year	HBPEs	2014	HBPEs/Hangzhou/ Zhejiang Province
	S13	Teaching Regulation for Hangzhou Wenlan Middle School	HWMS	n/b	HWMS/Hangzhou/ Zhejiang Province
	S14	The Assessment Policy and Report Card	WBS	n/b	WBS/Hangzhou/ Zhejiang Province