

LI, JIAYAO, Ph.D. What Do We Know about the Implementation of the Quality Rating and Improvement System?: A Cross-cultural Comparison in Three Countries. (2015)
Directed by Dr. Linda Hestenes. 248 pp.

Quality Rating and Improvement Systems (QRISs) provide an infrastructure that maintains as well as supports the enhancement of early childhood education (ECE) programs quality. It is being established worldwide. However, little is known about the implementation process and how this relates to implementation success. The purpose of this study was to have a better understanding of QRIS implementation through investigating QRIS in three regions within three countries (U.S., China, Singapore) based on implementation science framework (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), and cultural background of each country. A case study approach was used to examine the components of QRIS, implementation process, effect, and stage of each region's QRIS. Results showed that the components of QRIS in each region looked different, with U.S. (North Carolina) having the most comprehensive components, and China (Beijing) and Singapore having fewer, although each still had critical components within their systems. The three countries also possess different values, strategies, and resources to support the implementation of QRIS. Based on each region's implementation process, QRIS in North Carolina is in its full implementation stage with several major and minor adjustments going on; Beijing's QRIS is at a mixed stage with variations in the development of different drivers; Singapore's QRIS is clearly at an initial implementation stage. Implications for each region's QRIS are discussed under their unique cultural, political, and historical backgrounds within each country. Further implications for

implementation science framework as well as how it might be adjusted for each region's QRIS implementation is discussed at the end.

WHAT DO WE KNOW ABOUT THE IMPLEMENTATION OF THE QUALITY
RATING AND IMPROVEMENT SYSTEM?: A CROSS-CULTURAL
COMPARISON IN THREE COUNTRIES

by

Jiayao Li

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

Linda Hestenes
Committee Chair

© 2015 Jiayao Li

APPROVAL PAGE

This dissertation, written by Jiayao Li, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair Linda Hestenes

Committee Members Deborah Cassidy

Karen La Paro

Catherine Scott-Little

Date of Acceptance by Committee

Date of Final Oral Examination

ACKNOWLEDGMENTS

I must first thank my dissertation committee chair, professor Linda Hestenes. I would never be who I am now without Linda's generous support and step by step guidance since the first day we worked together. In my five and half years in graduate school, she provided thoughtful critiques that have strengthened my work, helped me with adjustment in a new country and culture, and encouraged me to be myself, as well as walk out of my comfort zone to accept challenges. I also want to thank the other members of my dissertation committee: professor Deborah Cassidy, Karen La Paro, and Catherine Scott-Little. I am amazed at their passion and great achievement in creating quality early care and education environments for all young children. Each of them taught me what it means to be a scholar, a teacher, and a better human being. Thank you for pushing me to be my best and providing encouragement along the way.

I also appreciate all the faculty members, staff, and graduate students in the department of Human Development and Family Studies for creating a supportive and inspiring environment. Special gratitude goes to professor Jonathan Tudge, who allowed me to do what I was interested in and to be who I wanted to be.

I would not have completed this study without the willingness of the early childhood teachers and directors, and government administrators who participated in the interviews and surveys. Thank you for your daily work with young children.

My friendships at UNCG have sustained me personally and professionally. In no particular order, I thank Alyson, Amy, Angel, Christine, Elisabeth, Hongjian, Jinni, Nan, Rebekah, Teressa, Yudan, and Yue.

I will never be able to thank my family adequately for their love and support. My mother and father encouraged me to pursue a life that I want to live. My other family members showed their biggest support in every domain. I am grateful for the opportunity to call you family. I also thank my honey, Zhiyuan, for supporting me through the process.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
 CHAPTER	
I. INTRODUCTION	1
II. INTRODUCTORY LITERATURE REVIEW	5
The Cultural and Political Context in Three Countries/Regions	5
North Carolina	6
Culture and Political Context of the U.S.	6
Development of QRIS in U.S. and in North Carolina	8
QRIS in North Carolina	9
People’s Republic of China (China)	12
Culture and Political Context of China	12
Development of Early Childhood Education in China	14
QRIS in Beijing.....	16
Singapore	18
Culture and Political Context of Singapore	18
History of Early Childhood Education in Singapore	18
QRIS in Singapore	19
III. IMPLEMENTATION SCIENCE AND LINKS TO QRIS	23
Definition and Model Description	24
Definition	24
Model Description	24
Utilization of Implementation Science	28
Applying Implementation Science to QRIS	30
Drivers.....	31
Stage of Implementation	33
Components of QRIS	37
Quality Standards.....	39
Quality Measurement and Rating	49
Quality Improvement	53
Financial Incentives	57
Consumer Education	60
Integrating QRIS Components under Implementation Science.....	63

Research Questions	65
IV. METHODOLOGY	66
Selection of Methodology	66
Procedure	68
Sampling	69
Interview Sampling	69
Survey Sampling	71
Position Statement	73
Measurement	75
Document Review	75
Survey	75
Interview Protocol	76
Reliability and Validity	78
Data Analysis Plan	78
V. RESULTS	81
North Carolina	81
Components of QRIS in North Carolina	81
QRIS Implementation in North Carolina	88
The Effect of QRIS Implementation in North Carolina	100
Implementation Stages of QRIS in North Carolina	115
Singapore	118
QRIS Components in Singapore	118
QRIS Implementation in Singapore	125
Implementation Effect of SPARK	137
Implementation Stages of SPARK	152
Beijing	156
Components of Beijing's QRIS	156
Consumer Education	161
QRIS's Implementation in Xicheng, Beijing	162
Implementation Effect of KQRS in Xicheng, Beijing	170
Implementation Stage of KQRS	179
VI. DISCUSSION	182
Integration of the Results	182
North Carolina Star Rated License System	182
Singapore Preschool Accreditation Framework (SPARK)	187
Beijing Kindergarten Quality Rating System (KQRS)	191

Contextual Comparisons across the Three Systems	194
Theme 1: Leadership.....	196
Theme 2: Organizational Climate and Communication Style.....	201
Theme 3: Performance Assessment	207
Feasibility of Implementation Science in the Three Countries.....	208
North Carolina Star Rated License System	209
SPARK.....	210
KQRS	211
Implications for the Original Framework	213
Future Directions	213
North Carolina Star Rated License System	213
Singapore Preschool Accreditation Framework	214
Beijing Kindergarten Quality Rating System	214
Contributions and Limitations	215
Contribution	215
Limitations	216
Future Directions for Research in QRIS	218
Conclusions.....	218
REFERENCES	220
APPENDIX A. SURVEY.....	235
APPENDIX B. ADMINISTRATOR INTERVIEW	242
APPENDIX C. CULTURAL CROSS-COMPARISONS.....	245

LIST OF TABLES

	Page
Table 1. Interview Program Information	70
Table 2. Teacher Demographics (Survey)	72
Table 3. Descriptive Survey Results	102
Table 4. Correlations between Teacher Satisfaction and Teacher's Feeling about QRIS (Singapore).....	106
Table 5. Discussion Themes	195

CHAPTER I

INTRODUCTION

The quality rating and improvement system (QRIS) is the system that aims for improving early childhood education (ECE) program quality to benefit children and their families through evidence-based activities and different kinds of support. The word “system” is the core concept of this term, which indicates that QRIS is not a one-time, single practice or intervention, rather it is an ongoing, multi-layered infrastructure that aligns each component to achieve multiple goals. In general, QRIS is composed of both a rating process and an improvement process. In the United States, a QRIS should be used for (a) increasing the access to high quality early childhood programs, (b) creating systematic improvement strategies to help programs improve, (c) providing resources to support the improvement and sustainability of high program quality, and (d) promoting public awareness, especially consumer’s awareness of early childhood program quality (National Association for the Education of Young Children [NAEYC], 2011).

As the implementation of QRIS grows across states and around the world, it has become an urgent issue to understand the mechanism of QRIS implementation. The translation from research to practice needs careful planning, close attention, and strong support from scientific studies (Halle, Metz, & Martinez-Beck, 2013). However the growth of information and evidence to support implementation lags behind the growth of

QRIS. Much more information is needed to have a more comprehensive understanding of the systems and how they are functioning.

A QRIS does not exist as an isolated entity. It emerges under the cultural and political context of a society, is largely affected by those backgrounds, and in turn, these influences shape the development of a society. For example, the launch of QRIS is partly in response to access inequity among families. It is affected by local population, education level, policies and regulations. However it also shapes how the society views ECE and how families are affected. A cross-cultural study may enable us to see how QRIS develops and also shapes the environment in which it is embedded, and in turn, may promote a more comprehensive understanding of QRIS implementation both now and in the future.

Other countries also have quality rating systems or quality rating and improvement systems, even though they may be called something different or may include more or fewer components in each system. Beijing, as the capital city of China launched a quality rating system in the late 1980s (DOE; Pan, Liu, & Lau, 2010). Now being guided by the *Guidelines for Kindergarten Education* (trial version; Ministry of Education in People's Republic of China, 2001), programs are evaluated to inform parents in their choice of ECE programs for their children. However, it appears that the measurement that was used for evaluating quality has not been validated and the efforts toward enhancing program quality are not as effective based on program directors, and teacher's perspectives (Liu, 2010). Singapore also has recently established the Singapore Preschool Accreditation Rating System (SPARK, 2010). Very limited information is

available in the effectiveness of this system at this point. Both systems may benefit from further exploration.

Implementation science is a promising theoretical framework successfully applied in nursing and medical studies, and is now being introduced into early childhood education field (Halle et al., 2013). This framework helps individuals transfer empirical research to real life practice in a systematic way. The current study aims to use implementation science as a guiding framework to investigate the development of QRISs in different countries and to link their implementation to each country's cultural and historical background.

To the researcher's knowledge, this study is the first of its kind to compare QRISs across cultures from an implementation perspective. To compare QRISs across the cultural and political context of three countries will add to the literature of QRISs worldwide. In addition, by learning and probing different QRISs under the guidance of implementation science, it is helpful for researchers to better understand how the components of the QRIS work, jointly as well as independently. Hopefully the work of this study can inform future direction for the QRIS in each country.

This study focuses on the implementation of QRIS in three countries. The second chapter will include an introductory literature review on the cultural background as well as the early childhood history in each country. In the third chapter an overview of implementation science will be presented along with an extensive literature review on the QRIS as it currently exists in each country, followed by a discussion of research gaps. In the fourth chapter, the methodology of the study is described along with the measures.

The results are presented in Chapter V, followed by discussion of the findings and conclusions in Chapter VI.

CHAPTER II

INTRODUCTORY LITERATURE REVIEW

The Cultural and Political Context in Three Countries/Regions

As the needs of non-parental care grow in most countries in the world, the need for early childhood education or care service grows simultaneously. With the growth in child development, brain, and psychology research, more and more people, including parents, policy makers, and researchers, are paying attention to early childhood education and recognizing the importance of early learning for individuals as well as for societies. The care and education that children receive from the non-parental settings become an equally important resource in their early learning.

Quality rating and improvement is the fundamental strategy that is used in many countries to guarantee the quality of early childhood education. While QRIS in the United States has a fair amount of documentation, research, and reports demonstrating what it is (Mitchell, 2009; Tietze, Cryer, Bairrão, Palacios, & Wetzel, 1996; Whitebook, Kipnis, Sakai, & Austin, 2012), less is known about what it looks like in the other two countries, including the components of QRIS, the implementer of QRIS, the use of the results of QRIS, and so on. It is important to understand how QRISs fit in different countries and how they make the system work to inform the development and implementation of high quality childcare. This does not imply that one has to copy whatever is successful in other countries or regions. A very straightforward goal of a cultural comparison study is to look

more broadly at the QRISs across countries and to make sense of each QRIS under certain cultural and political circumstances. Another expectation of a cross-cultural comparison study is that it may allow us to see the implementation path of QRIS, which in turn sheds light on the road ahead of the QRIS in each country or region.

North Carolina has been a leader in the implementation of the QRIS across the United States (Office of Planning, Research and Evaluation, 2010). Beijing is the first province in China that started the QRIS, and has been the leader of the QRIS in China (Pan et al., 2010). Singapore has been through a very intentional development process over the last five years to establish their QRIS. The three regions are comparable in their position in each country in terms of the development of QRIS. Before going into the discussion of QRIS of each country, it is fundamental to understand the cultural and political context of the three countries as well as the history and status of early childhood education in each of them. The purpose of this chapter is to set the groundwork for further discussion of more specific issues surrounding QRISs in these three countries.

North Carolina

Culture and Political Context of the U.S.

The United States is one of the most diverse countries in the world. The diversity refers to many aspects, including its multi-ethnic population, large variety of landscapes and resources, and different languages and cultures. In general the cultural context of the U.S. could be categorized as “western,” which has the typical characteristics such as individualism, egalitarianism, and faith in freedom and democracy (Gray, 2012). Located in the middle of two oceans as well as having a relatively short history, America had the

greatest flexibility in developing its own culture. The diversity of U.S. culture leads some researchers to categorize the culture as a salad bowl in which every subculture is distinct, yet connected with each other, in contrast to a melting pot, in which all the cultures are mixed and become one distinct culture (Adams, & Strother-Adams, 2001, as cited in Wikipedia). This cultural characteristic is reflected in acknowledging the differences and respecting each subculture in many fields, for example, in education.

In terms of political context, Daniel Elazar (1984; as cited in Gray, 2012) argued that the United States shares a general political culture which was composed of three subcultures: individualist, moralist, and traditionalist. The individualist political culture emphasizes capitalism, with the government playing a more limited role. Under such a political culture, the ultimate goal is to maximize individual benefits. The moralist political culture emphasizes commonwealth. The government plays a critical role in balancing benefits and maximizing public interest. The last political culture, traditionalist, aims to balance the social and economic hierarchy in the society. The three political cultures are spread unevenly in this nation. However, other people argued that the influence of contemporary culture has exceeded the influence of the historical cultures (Gray, 2012). The contemporary political culture features in its diversity in racial and ethnic groups. States vary their policies according to the heterogeneity of their populations (Gray, 2012). North Carolina has great diversity in terms of race and ethnicity. According to the 2013 Census, 71.7% of the population are White, including none-Hispanic White (64.4%) and Hispanic White (7.3%), 22.0% are Black or African American, 8.9% are Latino or Hispanic American, 2.6% are Asian American, 0.1% are

Native Hawaiian, and 2.0% are other race or multiracial American (2013 Census).

Policies are developed and implemented based on this diversity characteristic.

While the political ideology is important, public opinion is another important piece of political culture. Public opinion refers to the attitudes of individual citizens on public issues (Gray, 2012). One study found a positive relationship between public opinion and political philosophy, indicating that public opinion reflects political culture to some degree (Gray, 2012). It is evident that early childhood education is gaining more and more attention in this state. A recent poll (September, 2014) reported that more than 83% North Carolina voters believe that investing in early childhood education will benefit the economy in both the short and long run. The trend may be reflected in policymaking and subsequent political actions.

Development of QRIS in U.S. and in North Carolina

Early childhood education in the United States has been through a long evolution process. The formal childcare movement began in the 19th century due to the large requirements of women going to work and leaving little time for them to take care of their young children. At that time, childcare was available to families with more financial resources and higher socioeconomic status. Most of the money to support childcare came from private resources. It was not until the World War II when the nation had to fund childcare to support women who were supporting the war effort, that childcare finally became available to poor families. It marked the beginning of public support for young children (Pounder, 2010). In 1965, Head Start, a federally funded initiative, was launched

as a comprehensive preschool program to support poor children and their families in their emotional, social, psychological, and nutritional needs (Pounder, 2010; DCDEE).

The first statewide QRIS was created in 1998 in Oklahoma. Originally most systems started with a Quality Rating System, or QRS, but several states in the planning stages began to describe their systems as QRIS to include the two major purposes of these systems: Rating and improvement (Mitchell, 2009). Since then, QRIS has been used as a systematic approach to identify quality, and reward early childhood programs that provide high quality services. By 2010, 25 states had launched a QRIS. Twenty-one of the 25 statewide QRISs allowed school-age programs to participate and 22 allowed family child care homes to participate in their QRIS. Twenty of the 25 statewide QRIS allowed state-funded prekindergarten programs to participate, as well. In 2014, almost all the states either have launched a QRIS or are in some phase of QRIS development: Thirty-eight states have QRISs launched, two have regional QRISs launched, 6 are pilot states, 9 states are planning for a QRIS, and only one state requires their state legislature to pass a QRIS. Compared to 2009 when there were 20 statewide QRISs, and merely 5 in piloting phases, ECE has witnessed a dramatic increase of QRISs in the past five years. Now QRIS is moving to a new phase from emphasizing assessment and rating of childcare programs to supporting ongoing improvement of quality in early childcare settings (NAEYC Quality Rating and Improvement Systems Toolkit).

QRIS in North Carolina

In North Carolina, an estimated 64% of children under six years old have both parents working. This results in over 400,000 children under four in need of some child

care arrangement (U.S. Census Bureau, 2006, American Community Survey, as cited in North Carolina Child Care Resources & Referral Council, n.d.). Thus, the need for non-parental care is great in this state. The first childcare licensing law, which had only minimum safety and health requirements for ECE programs, was passed in 1971 in North Carolina (Pounder, 2010). Childcare teachers were required to be at least 16 years old, and were not required to have any education background. The first state government funded full-day kindergarten was launched two years later after the law passed. However, the first statewide initiative that focused on quality rating and improvement was not launched until 1986, more than a decade later after the state started to fund childcare programs. This rating system was an “A” and “AA” rating, and included the rating of teacher education and teacher-child ratios. The main purpose of the rating was to inform parental choice of childcare and also to start the process of childcare quality improvement based on the ratings (Pounder, 2010). However this rating was confusing for parents to understand since people tend to think an “A” represents the top rating, and were not aware of an “AA” rating (Pounder, 2010).

Even in the early 1990’s, North Carolina still had the worst early childhood education standards as well as a very high teacher turnover rate, and a very low record in SAT scores across the nation (Pounder, 2010). Upon seeing this, some education advocates and governors started to put actions into the improvement of quality education, including early childhood education. In 1993, the Smart Start bill was passed for the purpose of providing every young child in North Carolina the opportunity of accessing high quality childcare. By 1997, with the two-tiered rating system and Smart Start as well

as other supportive systems like the Teacher Education And Compensation Helps (T.E.A.C.H), the quality of childcare gradually improved, so did the long-term outcomes, such as SAT scores. In 1999 a new five-level Star Rated License for childcare facilities based on achieving higher standards was established (Pounder, 2010).

No one has explicitly and definitively defined what ECE quality is. The definition of ECE quality in the United States is influenced greatly by the National Association for the Education of Young Children (NAEYC) and what is known as developmentally appropriate practice (DAP). In the description of NAEYC, a high-quality program is described as providing “a safe, nurturing environment that promotes the physical, social, emotional and cognitive development of young children while responding to the needs of families” (as cited in Schuyler Center for Analysis and Advocacy, 2012, p. 2). They further clarify the components of a high-quality classroom should include: A developmentally appropriate curriculum, adequate teacher training, a safe environment, small group size and low adult to child ratio, and parent-teacher communication. This definition is apparently influenced and supported by evidence-based DAP.

States customize the toolkit that is used to evaluate quality according to the goals, political contexts, and regulations. The North Carolina Division of Child Development and Early Education (DCDEE) supervises and provides subsidies and financial incentives to childcare programs, in North Carolina. The reimbursement rate that corresponds to program quality depends not only on the quality of the program, but also the county where the program is located. Smart Start in North Carolina also provides resources and guidance for programs and professionals serving young children in the state. Other

organizations, such as Child Care Resource and Referral (CCR&R), and Teacher Education and Compensation Helps (T.E.A.C.H) provide technical assistance as well as resources to support teachers' professional development, program improvement, and parents' awareness of quality of childcare settings. Policies of each organization should not be in conflict with others but should align with and support the implementation of the others.

North Carolina has established a model of QRIS that supports programs, children, and families relatively well. However, even with a quite comprehensive development of QRIS in North Carolina, problems and questions remain. In a later section, the main problems that relate to the implementation of QRIS in the North Carolina as well as in the U.S. in general will be discussed. This relatively successful model, however, may not be copied in other countries directly. Grounded in specific cultural and political contexts, China and Singapore should develop their own QRIS (Hu & Li 2012).

People's Republic of China (China)

Culture and Political Context of China

There are 56 officially recognized ethnic groups in China. Although each of the ethnic groups keep some traditions and unique cultural characteristics, due to a long history of China, many cultures have merged into the mainstream culture, which is known as Han's culture. Han is the largest ethnic group in China (Census Bureau of China, 2010). Compared to the cultural context of the United States, Chinese culture is more like a "melting pot" in which subcultures merge with each other and form a rather uniform culture. The fact that China is governed by one political party also explains why

the cultures merge with each other instead of remaining distinct. This cultural characteristic determines a more single and standard policy that applies to most parts of the country. Take education as an example; schools in most provinces (states) use the same set of curriculum materials, and leave little flexibility to each province or school.

Most social values are derived from Confucianism and Taoism. Confucianism was the official philosophy in historical times, and is still influencing the ways people think and behave nowadays (Hu & Szente 2009; Moore & Morris, 1967). For example, hierarchical authority is one of the main principles of Confucianism. It emphasizes the respectfulness of authority figures and the compliance to elders or authorities (Fewsmith, 2010, Moore & Morris, 1967). However, since 1978 when the reform of the economy began and China widely opened the door to the world, western beliefs and ideologies flooded in this country through media, goods, and people. Although collectivism still dominates the cultural beliefs and behavior of people, individualization also emerged and has grown rapidly through the past decades (Wang & Zheng, 2013).

China is the most populous country in the world. The resources that are available to each individual are limited. There is a high level of economic inequity in this nation. China is one of the few countries that openly endorse a socialist perspective. The Chinese government is usually described as authoritarian, with heavy control in many areas, such as freedom in Internet access, freedom of the press, and freedom of assembly (Fewsmith, 2010). There are pros and cons in every political context. With a large population and limited resources, authoritarianism is helpful in maximizing the public benefit to promote

the common wealth. However, the restrictions on every domain make it hard to be creative to some degree (Wang & Zheng, 2013).

Beijing is capital city of China, and also the second largest Chinese city by urban population after Shanghai and is the nation's political, cultural, and educational center. Beijing Municipality currently comprises 16 administrative county-level subdivisions including 14 urban and suburban districts and two rural counties. In 2010, the annual education expenditure per secondary school student in Beijing was ¥20,023 (\$3000 USD), exceeding any other province or city in the nation. The cultural and political context in China as well as in Beijing determined the characteristics of early childhood education in this country while maintaining traditional beliefs and practices embedded in Chinese culture, such as collectivism and authoritarianism.

Development of Early Childhood Education in China

Early childhood education (ECE) in China is a part of the education system, and is supervised by the national education department. It is not a compulsory education, however, the public ECE is funded and managed by The Ministry of Education of the People's Republic of China. The other types of auspices include institutional ECE which are held by public institutions, like schools, companies, or community programs, as well as private ECE programs, which are operated by private businesses or specific individuals. There are also programs that are a mix of public and private institutions, which may be sponsored by national, institutional, or private organizations. Since ECE is a part of an education system, the transition from kindergarten (You'eryuan, which is

usually called ‘preschool’ in the U.S.) to primary school is greatly valued and is reflected in ECE policies.

The emergence of public early childhood education in China follows a similar track to that of the U.S. When the requirements of women in the workforce increased in the 19th century, the requirement for non-parental care increased as well. The first childcare in China appeared at the end of Qing dynasty (1930) when China was forced to open the door to the world, especially to the western world. At that time most of the structure and practices were imported and learned from Japan’s childcare system (Yang, 2004). However, unlike Japan, women did not share equal rights in education in China at that time and most of the caregivers in childcare were nannies, nurses, or maids (Yang, 2004). Thus, although the format or structure was like Japan, the quality of childcare was not comparable to western countries. Little improvement in quality was seen until the rights for women to receive education was officially established and protected by laws in the 20th century. There was a significant positive change in the quality of caregiving as well as education compared to previous decades.

Moving toward contemporary times, the ideology and practice in early childhood education became more and more similar and globalized in many different countries (Yang, 2004). This trend is partly based on the research in child development that was done through these years. People started to reflect on the traditional practices and make changes to it. The study that Tobin Wu, and Davidson (1989) did to compare preschools in Japan, China, and the U.S. presented many differences among three of them; however, the overall philosophy that was applied in the classroom of each country tended to be

similar. This globalizing trend was in fact a western value. This trend remains true in China nowadays.

QRIS in Beijing

The Kindergarten Quality Rating System (KQRS) was launched in 1989 in Beijing, China under the supervision of the Department of Education (DOE; Pan et al., 2010). Kindergarten in China refers to full-time programs serving children aged 3 to 6-years-old (Hu & Szente, 2009; Pan et al., 2010). Children who are under 3 are typically taken care of by their parents, grandparents, or in nursery schools, all of which are not under supervision of any accreditation system (Pan et al., 2010). By the time it was launched, all the childcare centers followed the *Kindergarten Work Regulations and Procedures* (as *Regulation*) which was issued by the National Education Committee of the People's Republic of China in 1989 as an effort to promote and measure curriculum reform in Early Childhood Education (Zhu & Zhang, 2008, as cited in Hu & Szente, 2009). It emphasized the following aspects: (a) child-initiated activities, (b) individual differences, (c) the importance of play, (d) an integrated curriculum, and (e) the process of activities (Zhu & Zhang, 2008).

The DOE monitors the quality of all types of kindergartens (i.e., public and private) in China. Every kindergarten needs to meet the national legal requirements to start operating. Beyond this level it is voluntary to apply for a rating. Based on national laws and regulations, each province (state) or city proposes a definite and detailed document that lists performance indicators for each level and category of quality. Thus the ranking standards vary across provinces or cities. For example, the Rank and

Category standards (R&C) of Beijing's KQRS refers to Rank (ji) and Category (lei).

Rank refers to the values, physical environment, staff, and management of a program;

Category refers to the quality of teaching, hygiene practices, teacher-child interaction,

and support of child development (Pan et al., 2010). Theoretically, by combining the

scores from Rank and Category there are 9 possible combinations of ranking. The top

ranking is R1C1. However, beyond R1C1 there is a Demonstration Level, which is

determined by a different set of criteria (the Beijing Demonstration Kindergarten

Selection Standards). The purpose of ranking is to (a) sustain staff and program quality;

(b) improve program quality; and (c) inform parents' choices (Liu & Pan, 2008)

In 2001, the DOE issued the latest document of *Guidelines for Kindergarten Education* (trial version; Ministry of Education in People's Republic of China, 2001).

This national curriculum guideline emphasizes more specific educational and

developmental goals for children. This guideline also links to Developmentally

Appropriate Practice (DAP) better than any other regulations. Each province then adjusts

their regulation criteria based on the *Guidelines for Kindergarten Education* (trial; as

Guidelines).

The definition of ECE quality in China can be retrieved from the requirements for

ECE programs in the *Guidelines*. In the *Guidelines* a high quality program is required to

be able to create a healthy environment with rich learning experiences for children.

Children should be able to learn through play, and the teacher needs to respect individual

differences to promote child development in all areas. In the requirements, the definition

of high quality ECE programs have the same features as that in the United States, such as

a healthy and safe environment and promoting holistic development in children.

However, Beijing, like many other provinces, is not using standardized rating scale for assessment. Even though China has a relatively long history of quality rating of ECE programs, it has inherited many challenges that are embedded in the nation's policy and regulations.

Singapore

Culture and Political Context of Singapore

Singapore is one of the world's major commercial centers, with the fourth-biggest financial center and one of the five biggest ports (Quah & Quah 1988). Its globalized and diversified economy depends heavily on trade, especially manufacturing. Singapore, like the U.S, embraces diverse subcultures, including different races, ethnicity groups, and immigrant families from other countries. In 2009, the government census reported that 74% of residents were of Chinese, 13.4% of Malay, and 9% of Indian descent.

Singapore's government is viewed as a hybrid of authoritarian and democratic. The Singaporean economy is known as one of the most free, innovative, competitive, and business-friendly around the world (Quah & Quah 1988).

History of Early Childhood Education in Singapore

In contrast to China's one-child policy, Singapore encourages each family to have more than 2 children if they can afford it. Based on a strong economy Singapore is able to invest a lot in education, including early childhood education. A good variety of subsidies are provided to childcare providers, working mothers, low-income families, and children

themselves. A total of Sg\$44,780,415 (approximately 30 million U.S. dollars) in child care subsidies was disbursed in the financial year of 1996 (Lim, 1998).

One challenge for early childhood educators or educators in general is to make children understand that they are equally identified by their nation even though they speak different languages and share different traditions (Honig & Lim 1998). Thus, there is an emphasis on improving their mother tongue, and also a focus on Asian values across subjects in their curriculum. In Singapore, childcare and kindergarten are viewed as valuable parts of education, even though they also have an emphasis on serving families. One of the requirements for teachers is to be bilingual, and they are usually required to speak English in class. However, the professional status of teachers are, like many other countries, not as desirable as it should be. According to a past survey of 28 teachers for 95 preschool children, only 11% of them have professional qualifications. Teacher experience varied from 1 year to 18 years (Lim, 1992, cited in Lim, 1998). The rapid expansion of care service and high demand on teacher qualification created more challenges for Singapore's early childhood education system. Childcare quality needed to be improved while balancing a low cost for low-income families. Advocates stressed the importance of improving quality based on stringent and strong evidence. Thus there was a call for developing an evaluation system for curricula for preschoolers and teachers/caregivers in Singapore (Lim, 1998).

QRIS in Singapore

Singapore is the country that has the shortest history of implementing a quality rating and improvement system among all three of the countries. In January 2011 the

Pre-school Accreditation Framework, *SPARK* was officially launched in Singapore (Anonymous, 2013; Wu, 2013). Its aim was to build a path for continuous improvement in teaching, administration, and management of early childhood education programs. There are five dimensions of core values that underpin *SPARK*. Children's development is the central value in this model; leadership is the vision that sets the tone for programs and manages challenges; Professionalism with Impact refers to the expectation of teacher as specialist with a strong pedagogical competence to create learning experiences for children; Innovation with Purpose requires programs to develop the sensibility and flexibility of change; Partnership for Growth asks programs to involve families and communities as their resource to enhance the holistic development of children (Wu, 2013).

The process of participating in evaluation is composed of four stages: Registration and regulation/licensing, self-appraisal, quality rating, and accreditation (Anonymous, 2014). The first stage is mandatory, in that every preschool or childcare center needs to meet certain criteria to be licensed for operation. From the second stage on the application is voluntary. In the second step programs will be asked to apply for a Quality Assurance Consultancy, with which programs self-appraise their quality based on a measurement scale and receive professional instructions on ways to make improvement before moving to the next stage. In the third stage, an outside expert will evaluate program quality based on the same measurement scale that was used in self-appraisal. At last, the Ministry of Education (MOE) will make decision on the program's quality and

accredit programs that meet the standards. This result will be posted on an official website. The accreditation is valid for three years (Wu, 2013).

The definition of quality in Singapore can be seen in the Quality Rating Scale (QRS; Early Childhood Development Agency, 2014), which is the only scale that is used to evaluate program quality, both in self-appraisal and by outsider evaluation.

Conceptually, the Quality Rating Model incorporates three parts: structure, process, and outcomes (QRS, 2010). Leadership is the driver that drives the structural components, which includes planning and administration, staff management, and resources. With these fundamental supports the process components, such as curriculum, pedagogy, and health and safety, can be promoted, which then lead to the holistic development of children. This model is aligned with the values that underpin *SPARK* well. The QRS measures seven domains of quality: leadership, planning and administration, staff management, resources, curriculum, pedagogy, and health, hygiene and safety. Each indicator in each domain is followed by detailed instructions on how to accurately interpret the item descriptions and provide explanations for abbreviations or new phrases.

Since this system is so fresh, no research on the effectiveness or problems of this system have been published. However it seems that it includes values that are crucial in program evaluation. For example, first, it emphasizes the central role of children's experience and development. Second, in the models leadership is considered as the driver of quality. This notion is also seen in the United States's QRIS, and is evidence-based. Finally, the structural quality serves as an infrastructure that incorporates both inside and

outside (e.g., families and communities) supports that will facilitate ongoing improvements.

The three countries in this study have distinct cultural and political backgrounds. Grounded on those contexts grows distinct early childhood education systems. To better serve children and families in each country, a QRIS should be fit for each ECE system. However, in reality it is difficult to align all aspects. Stakeholders and researchers need to work on promoting this alignment, and fitting the processes in order to maximize children's and families' benefits. In the following chapter the details of implementation science and each QRIS will be described and potential gaps will be identified.

CHAPTER III

IMPLEMENTATION SCIENCE AND LINKS TO QRIS

The main purpose of a Quality Rating and Improvement System (QRIS) is to create a strong infrastructure to support and sustain the quality of early childhood education (ECE) quality through multiple avenues. Other important purposes are to promote the awareness of components of quality for consumers (e.g., parents), and to provide incentives and resources to programs (NAEYC Quality Rating and Improvement System (QRIS) toolkit). Bearing with those purposes, the target groups of QRIS are children, their families, and ECE providers. Elements that are now included in a QRIS in most states are standards, accountability, program and provider outreach and support, financial incentives linked to compliance with standards, and consumer education (NAEYC Quality Rating and Improvement System (QRIS) toolkit). Administrators of a QRIS aim to build a system that incorporates key structures that will help programs acquire resources and promote awareness, with the ultimate goal being to serve children and their families in a quality way. Implementation science (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), a theoretical framework that has been successfully applied in the medical health domain, will be used to guide the discussion of different aspects of QRIS and to propose potential missing pieces that should be included in a QRIS. In this chapter, the model of implementation science will be described in general first and then features that are related to QRIS will be discussed in detail.

Definition and Model Description

Definition

The science of implementation is the study of the process of implementing intervention programs and practices that have research-based evidence. Implementation science is the way to replicate these evidence-based practices and to bring them to scale (Halle et al., 2013). “It is not the act of validating a program as evidence-based, instead implementation science is the study of how a practice that is evidence-based or evidence-informed gets translated to different contexts in the ‘real world’” (Halle et al., 2013; p. xix). Scaling-up, which is to develop an intervention from serving a small group of people to serve a larger population, is a further step to carry implementation science to a broad real world.

Model Description

There are different models of implementation science. However the model that is mostly used in ECE studies is by Fixsen et al. (2005). Compared to other models or frameworks, Fixsen et al. (2005) emphasizes the multi-level influences on implementation, from external influences to organizational and core implementation process components, which include the central role of the individuals who coach and train prospective practitioners and the practitioners themselves. The multi-level framework and the central role of process components make this framework a good fit for research in the ECE field since children are often thought of at the center of the model with schools, families, communities, and the whole society are surrounding them to provide optimal opportunities for child development. The model developed by Fixsen et

al. (2005) should be a good fit for ECE studies which involve multi-layers of an ecological system (e.g., school, home, individual). Thus, this model may provide guidance for practitioners in the ECE field to line up components in all ecological layers and to exerting successful interventions.

The model is composed of 3 drivers that are the core components of the infrastructure supporting practice, organizational, and systematical changes (Metz, Halle, Baartley, & Blasberg, 2013). The competency driver aims at promoting the confidence and skills of ECE educators in engaging in new programs and interventions. This driver is built through three steps of staff management processes: selection, training, and coaching. For example, a program director needs to set up systems for hiring, training, and mentoring all the teachers in the program. This driver also applies to infrastructures that serve the implementation of an intervention. If an administrator is implementing the intervention then he/she needs to, for example, set up a system of hiring staff. One last step, sustaining the competence through ongoing performance assessment, is not explicitly noted in the model, despite its importance in maintaining the effect of implementation along all the stages.

The second driver is the organization driver which aims to “create and sustain hospitable organizational and systems environments for effective services” (Metz et al., 2013, p. 31). The organization driver creates an environment that supports and facilitates the competency driver. The three components of the organizational driver are systems intervention, facilitative administration, and decision-support data system. Systems interventions are the strategies for leaders and staff to work with external systems to

assure the alignment of standards, and the availability of all kinds of resources. For example, an intervention in education needs to have policies that apply to the specific intervention and policies that apply to education overall aligned to create an optimal outcome. Leaders should also pay attention to outreach resources, such as financial support to undergird a successful intervention. Facilitative administration utilizes data input to inform decision making around culture and climate of the organization to assure the success of implementation. Decision-support data systems provide data for decision making, including fidelity information and the correlation between fidelity and outcomes. The data system would support evidence-based decision making by providing data of the cost and the product to inform the leaders and staff how effective the intervention is and to make changes to make it more effective.

Lastly, the leadership dimension, includes a technical driver and an adaptive driver (Metz et al., 2013). The term “leader” can be referred to at every level, including the QRIS administration level, and implementation level (Damschroder et al., 2009). The leadership driver, including the technical driver and adaptive driver, is necessary when problems emerge. The technical driver is used when the problem is clearly defined, and the solution is agreed upon. The adaptive driver is used when the problem is unclear, and problem solving involves changes in philosophy, values, and beliefs. For example, when the policy of the intervention has a major conflict with the broader policy that overlooks the whole educational system, there needs to be a big change in the intervention policy to adapt to the broader policy or value. Leaders must facilitate aligning components of an intervention or a system. It is critical to keep a big picture in mind, and ensure that

processes and structures are in place to support successful implementation. Leaders should also always keep an eye on resources that could be made use of to provide support to the system (Paulsell, Tout, & Maxwell, 2013). Leaders are the key people to make the three drivers work. The three drivers of implementation science should align with and collaborate with and complement each other. However any implementation cannot be completed in one day. Each driver of the implementation moves from stage to stage.

There are four implementation stages in the model. The four stages are: Exploration, installation, initial implementation, and full implementation (Metz et al., 2013). An exploration stage is characterized as collecting information from all the sources that are available to form an initial idea of how to establish the system and how to make it work. An installation state is featured with lots of trial and error in the process of actualizing the initial idea that forms in the exploration stage. An initial implementation stage is a relatively stable, yet still under an experimental phase, which is seen as the stage to promote buy-in for all stakeholders, and to use data to improve the system. The last stage, full implementation stage, is the mature phase of an implementation. It is essential to sustain the effects from earlier stages and to maintain the support that enables its success. It usually takes two to four years to reach the full implementation stage. Sustainability is not a discrete stage but is considered a part of every stage. Evidence for each stage will be discussed in relation to specific QRIS components in a later section.

Utilization of Implementation Science

Only recently has implementation science gained widespread attention in the fields of education. The initial stage is usually the implementation of an intervention in a program, which is usually followed by an adjustment and adaption of the practice in the program. In this process, implementation fidelity (i.e., the support that ensures the intervention is implemented) as well as intervention fidelity (e.g., intervention dosage and quality) is documented through data collection and analyses (Downer, 2013).

Sustainability of successful intervention is usually achieved after the implementation is well adapted and monitored along the road. The actualization of implementation science needs careful planning, many rounds of refinement, and strong leadership. Before discussing how to use implementation science to guide QRIS, a very complex system, it may be helpful to look at some examples of applying implementation science to relative single-dimensional interventions.

Literature has documented the utilization of implementation science in the educational field. One example comes from the Head Start CARES Demonstration project (Lloyd, Supplee, & Mattera, 2013). Head Start CARES is a large-scale national research demonstration designed to test the effectiveness of three social-emotional programs in Head Start settings. Specifically, this study examined the effectiveness of scaling-up the effective intervention in diverse settings. By using fidelity logs, researchers kept monitoring the scale-up process. These logs provided a place for teachers and coaches to report the implementation process by recording two components of fidelity: Intervention fidelity and implementation fidelity. Intervention fidelity refers to

“the extent to which core program components are delivered as intended” (Lloyd et al., 2013, p. 158). Implementation fidelity refers to the extent to which drivers are functioning as intended (Lloyd et al., 2013). Setting a threshold for fidelity set a standard level of implementation that all interventions had to meet. Thus, the implementation of the intervention was monitored with the threshold marked each time through the intervention year. A threshold is like a goal for each time. After implementing for a certain period of time, the team will need to evaluate the data that were collected along the implementation process. The threshold can be multidimensional to indicate the completion of the goal from different angles. However, challenges of developing fidelity measurement also existed. First of all, the definition of fidelity needs to be clearly and explicitly articulated. This is challenging because it is hard to reach agreement with all stakeholders. The second challenge lies in the design of measurement. The measurement needs to be reliable and validate itself. The last challenge is the disagreement in interpreting indicators of fidelity between teachers and coaches. Despite of the challenges, the authors acknowledged that fidelity instruments can help practitioners make decisions about the feasibility of potential adaptations and inform future directions. Frameworks or models of implementation science may provide a systematic process of how to scale up an intervention.

In this example, fidelity testing is a way to collect data about the effectiveness of the interventions. It is a part of the organizational driver, which provides support for further decision making as well as staff professional development in this system. The definition and threshold of fidelity should also be determined based on the standards that

are clearly lined up with research findings, as well as policies, and higher-level standards. Effective communication between coaches and teachers should be based on careful selection and effective training of the coaches. In solving the challenges of fidelity as well as measurement validity as mentioned above, a strong leader will help make appropriate and necessary changes on the content or the structure of the intervention. For example, if an educational intervention is proven overtime to be ineffective, or not as effective as it is supposed to be, then a change either in the way that measures its effectiveness or in the intervention design itself needs to be made. This Head Start CARES Demonstration project may give a snapshot of the mechanism through which the three drivers of implementation science work, especially how the organization driver and leadership driver work. Staff training is essential to make sure each evaluation and implementation of practice is accountable. While QRIS is a multi-dimensional system, it shares similarities with implementing a single-dimensional intervention.

Applying Implementation Science to QRIS

QRIS is not a single intervention practice or program like what was described in the Head Start project. It is a series of practices and processes that aim for changes in programs and systems. QRIS is composed of five components: quality standards, assigning ratings, financial incentives, and marketing and consumer education (Tout, Starr, & Soli, 2010). Quality standards refer to the quality that QRIS aims to promote. Alignment between the measurement and standards is the key for this component. In addition, it is crucial that the standards used in QRIS are evidence-based. The second component, assigning ratings, refers to the process of assessing quality based on the

standards through a variety of ways, such as observation, interview, and document review. The central challenge for this component is to establish strong reliability and validity for each measurement. The third component, quality improvement, refers to the strategies aligning the program's observed quality with the standards, and providing different kinds of support to help the program move forward. It is important to note that improvement should not be a one-time action, but rather an ongoing process that continuously pushes programs to progress. Financial incentives, the fourth component, is the way to support or reward quality improvement. Even though numerous programs may have benefited from the support, little research has been done to inform the appropriate amount and form of financial support in different situations and the effectiveness of such support (Metz et al., 2013). Marketing and consumer education, the last component, is the effort to reach the external world outside of ECE to promote understanding of quality in public, especially among parents. No research has been located regarding the effectiveness of different marketing strategies either.

Drivers

In order to apply implementation science to QRIS, one needs to look for ways to support the successful implementation of the five components based on the three drivers in the implementation science framework. First the competency driver aims to promote the skills and confidence in infrastructure staff. When this applies to QRIS, the staff infrastructure organization is the focus of the implementation. The first step is to identify the qualifications that would be required in data collection (e.g., observations, interview, and document review). Staff selection should be based on the qualifications and

dispositions to conduct observations, interview, and document review. The characteristics of a qualified staff may include the ability and willingness to adhere to a specific protocol (Halle et al., 2013), and dedication to details. Specific training is needed for conducting reliable assessments using observation measures. Every potential observer needs to achieve a certain level of reliability, which is usually 85% agreement with the gold-standard observer, before they begin to conduct observations for rating purposes (Tout et al., 2010). The performance assessment is also important. Along the process of assessing, observers should monitor their reliability regularly to make sure the rating is consistent. The term “staff” and “leader” refers to every level, such as the administration level, implementation level, and the program level. It is important to be clear that in the discussion of the current study, the leaders and staff refer specifically to people who are in charge of or implementing quality assessment and improvement, because they have the most direct influence on the implementation of QRIS.

The organization driver further supports the successful implementation of the competency driver. The decision support data system is crucial in providing QRIS with evidence to support decision making, such as the length of the observation, the dimension of quality that needs to be measured, cut-points of different rating levels, and the form and amount of financial incentives. Paulsell et al. (2013) argued that when merging quality rating data with data from the child care subsidy system, reimbursement rates need to be adjusted and applied to each child. Facilitative administrative support is the effort to align quality measurement with early standards, and align QRIS structure with policies. For the leadership driver, the leaders at the state level need to facilitate training

and provide structure for the timing of these activities. Leaders at the agency level must specify the schedule for raters and be clear about everyone's responsibility. In all, in order to implement every step effectively facilitative administrative support needs to plan out every detail on how to implement it. In addition, this system also builds a communication platform among stakeholders, like policy makers, practitioners, and parents. At last, the systems interventions require QRIS to work with agencies outside of the main system. Partnership with other organizations, such as the developer of the observational measurement, gold-standard raters, technical assistance, and financial providers, enables effective implementation. The drivers in the implementation of QRIS are based on an infrastructure that provide staff selection and training, actively aligns standards and policies, creates databases and develops strong leadership.

Stage of Implementation

Since QRIS is not a single-dimension intervention, it is important to note that each component may be at different stages at different times. For example, competency driver may be at the full implementation stage with complete structure of staff selection and training successfully implemented. It could be that system intervention is at an innovation stage with some partnerships being established and some being abandoned. However, the decision support data system may still be at the installation stage where much more research is needed to provide more solid evidence to support decisions. It is also important to consider that each component is intertwined with each other; it is impossible for one component to be fully implemented while other components move forward. For example, even though the implementation of financial incentives is part of

systems interventions, and has been implemented according to a certain criterion for a long time, the underdevelopment of decision support data system may prevent it from moving to another level due to the potential inappropriateness of a financial incentive criterion.

Financial incentives are used to support program improvement and reward improved programs. However, little research is available to decide the form and amount of financial incentives. Little is known about whether financial incentives are effective, how to make them effective, and the criteria used to decide different levels of financial support. Since implementation science is so new, more research is needed to document the process of QRIS as a systematic intervention. The research should examine staff selection, staff training, staff composition, the use of financial incentives, the administration process, and partnership with outside agencies. The recording of the entire process as well as the effects will inform further implementation and provide important information on how the current system works.

Evidence of the stage of implementation. The evidence of each implementation stage will help to identify the stage and guide future plans. The exploration stage is characterized by frequent, iterative interactions among and between stakeholder groups, knowledgeable experts, and perhaps other sites or organizations that have already implemented the new way of work. Evidence-based practices need to be selected and refined further at this stage. There are more questions than answers in this stage. Questions need to be asked and trials need to be taken in several domains, such as what measurement will be used, what financial incentives are appropriate for programs,

families, and individuals, what technical assistance may be effective in helping programs make improvements, how to set up a data system to be used in future decision making, and how to make the information not only accessible to consumers but also better understood by consumers. These questions will help form a blueprint of the new system. This stage is usually based on expert panel discussions, interviews, and document review (Metz et al., 2013).

The installation stage is characterized by altering, adjusting, or changing the original structure in order to initiate the new system. It is also the time to make decisions on financial incentives, including resources the incentives come from, the mechanism of rewarding, and the magnitude of incentives (Metz et al., 2013). An equally important step is to establish data systems to support decision-making in the future, such as deciding what evidence is required to send out technical assistance, and how many financial incentives are appropriate to help with improvement. At last, the preparation should also include consumer education. For example, at the installation stage, practitioners need to create paths of accessing information for parents, making sure childcare resource and referral agencies are in place to provide supports, and provide a budget for consumer education. In the installation stage, trial and error are not only allowed but are encouraged because it will help refine each component of the system. Careful staff selection, training, and coaching are to be made at the installation stage also. One of the key components of this stage is developing the competence and confidence of ECE practitioners. In addition, it is also a critical stage for the organization climate to form. A hospitable environment is based on similar beliefs, values, and strong supporting systems (Metz et al., 2013).

The third stage is the initial implementation stage. After the installation of the new system, the key activities of the initial implementation stage involve ways to maintain and promote continuous improvement based on a rapid-cycle of feedback from the established data system, and appropriate problem solving. For example, when technical assistance is found to be not very effective according to the evidence shown in the data system, a rapid adjustment needs to be made. Although the leadership driver is everywhere in the system, when problems emerge it is the time when the leadership becomes extremely important in decision-making and action taking. A continuous buy-in among stakeholders is also a key feature of the initial implementation stage. It is important to keep communication open to groups and individuals in order to help maintain an optimistic but realistic perspective.

The last stage, full implementation is signified when the new learning becomes integrated into practice at all levels: the individual, organizational, and systems levels. Evidence for this stage might include the degree of fidelity, the proportion of staff that are trained, and comprehensive functioning of the infrastructure and system components or the amount of time that has elapsed since the new program or practice was initiated (Metz et al., 2013). To sustain full implementation requires consistent and reliable support.

Sustainability is not a separate stage, rather it should be maintained within each stage starting from the initial stage and carried on across stages. Sustainability refers to maintaining the same high level of implementation quality at each stage. In order to be sustainable the supports for implementation should be consistent, adequate, and reliable.

The supports usually refer to financial and programmatic sustainability. For financial sustainability it requires that the funding stream is established and provides adequate funding for implementation with a high level of fidelity. On the programmatic side, supports such as policies, staff training, coaching, and performance assessment should be in place to inform data-driven decisions. It is also important to note that staff training or coaching is not a static, one-time event that happens at a certain stage. The continuous professional development for staff is an important prerequisite in guaranteeing sustainability within and across stages.

Implementation science will provide a theoretical framework to investigate the components of each QRIS. Even though the specific component may differ among the three countries, the three drivers and the stage evidence may hold true across cultures. The purpose of the study is to understand how to be successful in implementation under a given cultural background.

Components of QRIS

The components of QRIS are typically thought of as five critical components in the U.S. As mentioned previously, the five components are: Quality standards, quality rating, quality improvement, financial incentives, and consumer education (Tout et al., 2010). Each QRIS should develop quality standards at the beginning of their systems development. The standards, which may vary across different countries or regions, generally describe the goals for high quality, and are usually aligned with policies and scientific research findings. It is the foundation for the Quality Rating System (Tout et al., 2010). The quality rating usually involves the measurement used in the assessment,

the staff training set up to carry out the assessment, and the data management to inform improvement. Quality improvement usually includes the staff training which underlies the technical assistance to help programs improve, and the data management system designed to make use of assessment results. Financial incentives generally include the distribution of financial resources for quality improvement, the mechanism and the magnitude of the incentives, and the estimation of cost and effect (Mitchell, 2012). Finally, consumer education involves the dissemination of information regarding program quality, developing supportive institutions for consumers, and creating the communication path to consumers about financial incentives. The primary question of interest for QRISs is whether each key component is valid in making the differences it aims to make. This is also referred as intervention fidelity. After years of implementation, researchers and practitioners have acquired some evidence of the effectiveness of some of the components, with the goal of trying to make each of them more effective (Hestenes et al., 2014).

Even though there might be more or less components in QRIS in other countries, since the U.S. has the richest and most mature documentation of the implementation of the system among the three, the following discussion will be organized around the five critical components of a QRIS for each of the three countries of interest for this study. The ultimate goal, however, is to look at how all the components function jointly to be effective within each country.

Quality Standards

The United States. The Compendium of QRS and Evaluations (Tout et al., 2010) identified 13 categories that capture the range of standards used by QRIS to define quality within the rating structure. These include: licensing compliance (22), environment (21), staff qualifications (22), family partnership (21), and accreditation (19), administration and management (16). The remaining categories are included in nine or fewer QRSs: curriculum (9), ratio and group size (5) and child assessment (8), health and safety (4), cultural and linguistic diversity (2), provisions for children with special needs (6) and community involvement (6). Among the 13 categories, five of them were included in most of the states that had QRIS launched by 2012 (26). They are licensing compliance (26), environment (24), staff qualifications (26), family partnership (24), administration and management (23) and accreditation (21, Tout et al., 2010).

Developmentally appropriate standards. North Carolina (Office of Planning, Research and Evaluation, 2010) includes 7 indicators as standards. They are: licensing compliance, adult-child ratio and group size, curriculum, environment, staff qualification, family partnership, and administration and management. Licensing compliance examines if program had achieved minimum requirement of health and safety regulations (North Carolina state documents regulating health and safety in child care). After 6 months upon receiving the license (i.e., one star) programs can apply to be assessed to earn more stars (Office of Planning, Research and Evaluation, 2010). For the adult-child ratio and group size, it is stringently required to keep group size under a certain number for different age groups. For example, in infant classrooms (0-12 months), the required adult-child ratio is

1:5, and the maximum number of children in the room is 10. For 4-5-year-old children, the adult-child ratio should be under 1:13, and the maximum number of children is 25. The smaller the ratio or group size the more points a classroom will earn. The maximum possible point total is 7.

North Carolina requires programs use a developmentally appropriate curriculum that addresses five domains of development in teaching (Office of Planning, Research and Evaluation, 2010). There is a list that specifies four and five curricula for infant/toddler and preschool classroom respectively. The classroom environment is measured using an age-appropriate Environment Rating Scale (ERS; Early Childhood Environment Rating Scale Revised; Harms, Clifford, & Cryer, 2005; Infant/Toddler Environment Rating Scale Revised; Harms, Cryer, & Clifford, 2003; School Age Environment Rating Scale, Harms, 2013, and the Family Child Care Environment Rating Scale, Harms, Cryer, & Clifford, 2007), which will be discussed later in the quality rating section. Staff qualification applies to both directors and teachers, and includes education and experiences. The maximum possible points is 7. Take director qualification as an example, with a level I North Carolina Early Childhood Administrator Credential (NCECAC) or equivalent and two years of full-time verifiable ECE experience or one year experience in childcare administration, the director could receive 2 points. In order to receive 7 points, the director must have a level III NCECAC or equivalent and at least 4 years full-time work experience, either as a teacher or as an administrator (Office of Planning, Research and Evaluation, 2010). Family partnership examines if programs provide family involvement with systematic support, such as offering newsletters,

organizing parent advisory board, and parent information meetings regularly. At last, administration and management concentrate on program policy addressing staff professional development, staff benefit, and program regulation routines.

Quality standards in different countries may vary according to cultural beliefs, historical influences, or political context. There is a trend that, in early childhood education (ECE) field, beliefs and practices are getting more and more similar across countries, specifically, the western value is prevailing in the world (Pei, 2010). The so-called “western value” actually refers to rationalism and scientific value that leads to accumulated empirical research in child development and family studies, and in return, guides practices (Pei, 2010). Early childhood education in the U.S. is the best representative of using evidence-based practice to define quality.

Evidence-based standards. Evidence-based standards refer to the guiding principles developed from scientific research (e.g., Developmentally Appropriate Practice from NAEYC). Numerous studies show the relationship between several indicators and children’s outcomes (Hestenes et al., 2014; Sabol & Pianta, 2014). Teacher education or qualification, and training, teacher-child interaction, adult-child ratio and group size, and learning environment and instructional practice are the most studied structural and process indicators that are associated with child outcomes (Connor, Son, Hindman, & Morrison, 2005).

Teacher education has shown direct and indirect effects on child outcomes (Connor et al., 2005; Whitebook, 2003). Teachers with a bachelor degree and equipped with specialized training are found to provide the best preparation for children to the next

level (Whitebook, 2003). Teacher education is positively related to teacher's warmth, which in turn, relates to children's vocabulary skills (Connor et al., 2005). Less evidence is found on the impact of teacher experience in ECE field. There is, however, one study in primary school that showed teacher's experience did matter for enhancing their effectiveness (Clotfelter, Ladd, & Vigdor, 2007). Other evidence showed that teacher experience does not matter or has a modest positive effect, or has negative effect (Connor et al., 2005, Pianta, La Paro, Payne, Cox, & Bradley, 2002). Despite the ambivalent findings of teacher experience, it remains an indicator in many quality standards.

Teacher-child interaction and teacher-child relationships are two key factors that affect classroom quality. Findings for teacher-child interaction showed a positive relationship between different dimensions of teacher-child interaction and child outcomes (Burchinal et al., 2008). Teacher-child relationship also has proven to be a significant indicator in predicting child outcomes, such as children's perception of others and themselves, and their adjustment in future life (Colwell & Lindsey, 2003; Pianta, Hamre, & Stuhlman, 2003). Despite the importance of teacher-child interaction and relationship, there are not many indicators that address this dimension of quality in North Carolina's quality standards. One way to emphasize this indicator is to measure it as part of the rating. This could be a next step for policy maker to consider. In fact a QRIS Advisory Committee who met regularly from 2009 to 2012 is considering developing a new measurement to use in conjunction with or replace the widely used ERS (NC QRIS Advisory Committee, 2012, as cited in Hestenes et al., 2014).

Child-to-teacher ratio is another crucial element that influences the quality of education and caring (NICHD Early Child Care Research Network, 2000; Pianta et al., 2002; Smith, McMillan, Kennedy, & Ratcliffe, 1988). Another study found that teacher-child ratio alone is not predicting classroom quality (Pianta et al., 2005), rather program and teacher attributions have a more direct link with quality. However, teacher-child ratio may have an indirect effect on child outcomes. For example, Smith et al. (1988) found that adding a third teacher in the classroom had positive effect on teacher-teacher interaction, and teachers were also more involved in children's play, and had more communication with parents. It makes more sense not to drop teacher-child ratio as an indicator of quality, and at the same time, thinking about ways to make good use of a high teacher-child ratio to actually benefit children and the families is important.

With a high teacher turnover rate in ECE programs, increasing attention is being paid to teacher's development, including professional development and wage rate (Whitebook, Sakai, & Howes, 2004). One study found that programs that provide higher wages for teachers showed more sustainability in quality and had lower teacher turnover rates (Whitebook et al., 2004). However, research also found that preschool teachers do not leave just because of minimum pay. The other most mentioned reasons included high demand levels of classroom responsibility and relationship tension with coworkers or with directors (Wells, 2014). The study further found that it is the accumulated effect that drove teachers to leave a program (Wells, 2014). Thus, the examination of administration and management is the initial effort to make sure that teachers are provided with opportunities in professional development and receiving benefits for their work.

The establishment of the Teacher Education and Compensation Helps (T.E.A.C.H) was based on research that teachers' education and professional knowledge and wages matter in promoting optimal child outcomes. However more studies are in need to understand the relationship between administration indicators and program quality. Threshold studies that look at the minimum amount of supports (e.g., wage, professional development) may be useful in the future to inform policy decisions.

What do not appear to be included in North Carolina's quality standards are health and safety, child assessment, cultural/linguistic diversity, accreditation, and community involvement. Several of these indicators are required in the state licensing regulation (e.g., health and safety), and should be addressed before a quality rating launches. North Carolina is one of the three states that currently do not link QRIS to Accreditation (NAEYC Public Policy Fact Sheet, 2010). Less is known about the rationale for this decision.

The rapid growth and expansion of QRIS across states creates a shortage in research evidence to support each step in this complex system (Kirby, Caronongan, Malone, & Boller, 2014). With some ambivalent findings the quality standards need to be selected cautiously. Hestenes et al. (2014) found that the North Carolina star rated licensing system failed to distinguish program quality at each rating level, even though it does differentiate higher (i.e., 4 and 5 stars) and lower (i.e., 1 to 3 stars) quality programs. It appears that North Carolina has made use of many quality standards in their QRIS but some aspects appear to be missing. The current project will look more closely at the existing standards.

China. The early childhood education policy in China can be divided into three phases. The first phase is from 1950 to 1980. The *Outline for Kindergarten Education (Trial, as Outline)* and The *Kindergarten Regulation (Trial)* were released and used to guide early childhood education practices. The second phase was when the updated *Outline* was released and replaced the old *Outline* from 1981 to 2000. The third phase is from 2001 to present when the *Outline* was updated to the *Guideline for Kindergarten Education (Trial, Guideline)*. All the policies are used nationwide. In the first two phases, there were very clear-cut curriculum requirements in the *Outline* for children. The early childhood curriculum was closely linked to curriculum in primary school, such as literacy, calculation, physical education, moral education, music, and arts.

In the latest *Guideline*, however, the clear-cut curriculum in the *Outline* is replaced by more general expectations in five domains, namely, health, language, social, science, and art. Within each domain there are less specified requirements, and there is some emphases on family involvement in promoting health practice and independence of children (*Guideline*). These changes reflect the change in beliefs and values in ECE in China. The evaluation and improvement of ECE quality is guided by the *Guideline*. Since it is more a guidance of value and beliefs than guidance for practice, each province still has great flexibility in making their own policy and standards for quality fit with the local context. However, the resources for public kindergarten are mostly from the national level. Thus, theoretically, kindergartens in China are supervised by the Ministry of Education of China and need to comply with national policy before complying with any local policies. But the implementation of the policy and standards are at the local level.

The holistic child development requirement is seen across all three phases. However, the definition of holistic development varies along the timeline. Education in China is largely affected by politics. Policies need to comply with the spirit of Communism of China. Moral education is seen in ECE in the earliest *Outline*. However, the latest *Guideline* puts much less emphasis on moral education as well as linking ECE to the communist spirit of the country. More emphases is now put on identifying individual needs and using knowledge in child development to promote children's development in all areas.

As aforementioned, there is a trend of globalization of beliefs and practices in early childhood education. Chinese educators, researchers, and parents are being influenced by globalization in ECE values and practices (Pei, 2010). However, as reported by Pei (2010) traditional values are still dominant, at least according to parent interviews. However, very little evidence was found to support whether following western value or keeping with the traditional value is better for children who are born and brought up under the Chinese cultural and political context. In fact, people tend to agree that the globalization trend has both pros and cons. On the positive side, children are more and more viewed as independent individuals that bear differences in ability, interest, and personalities. Play is widely accepted as the primary way of young children's learning. The disadvantage of globalization is the disconnection with domestic culture (Pei, 2010). Most of the values or practices are derived from the findings from western world, in a certain population—usually middle-class White (Ogbu, 1988). It is

not clear if the findings apply to other countries with different cultural and political contexts.

There are, however, few studies which have looked at teacher-ratio, teacher professional development, teacher-child interaction, and teacher wages (Sha & Yao, 2014; Liu & Liu, 2007). However, most of the studies provided an overview of research related to a certain topic in the world, or looked at the phenomenon (i.e., teacher-child ratio, teacher-child interaction, teacher professional development, teacher wages) itself without linking it to classroom quality, or child outcomes. Only the study by Liu (2007) explicitly examined the relationship between group size and classroom quality. Through qualitative methods they found that when the group size was small teachers' self-reflection on their teaching focused more on meeting individual needs than on achieving goals for teaching.

In 2013, the Ministry of Education published the *Kindergarten Staff Ratio Standard (Temporary)*. In the document, for example, a 3-4-year-old classroom could have 25 children maximum, and 2 teachers and 1 nursery governess. The requirement for different age groups is slightly different. There is no detailed description of how the standards were adopted, or what evidence supports it. It is highly possible that this standard has no empirical research to support it and is mainly based on a global trend. Policy makers of ECE quality standards in China seem to be trying hard to align standards with research evidence. However, it is dangerous to use research evidence from elsewhere directly without examining its adaptation in Chinese cultural context.

Singapore. Singapore is a compact city-state that has rich cultural and racial diversity (Choo, 2010). Due to the unique features of the country, there is no difference

between national and local policy. Early childhood education is not a part of the educational system in Singapore. It is supervised and regulated by the Ministry of Education (MOE) and the Ministry of Community Development, Youth and Sports (MCYS) jointly (Ang, 2012). Teacher qualification requirements have been through a changing process from very minimum requirements to completing at least 5 “O” level credits including a passing of English. (An “O” level credit is acquired after four years of express education or five years of normal academic secondary education.) An effort towards developing professional pathways for teachers includes the Fundamental Course for childcare assistants, the Intermediate Course for childcare teachers, and the Advanced Course for center supervisors. This has recently been raised to a Diploma in Early Childhood Care and Education for teachers and another Diploma in leadership for program directors (Choo, 2010).

It was not until 2010 when the MOE introduced the Singapore Preschool Accreditation Framework (SPARK; Ang, 2012) which is the quality rating system for preschools, including childcare and kindergarten. The core standard, as introduced in an earlier chapter, is to promote holistic development of children. The main function of SPARK is to provide each program director with a reliable assessment in order to make a successful teaching, learning, administration, and management process (SPARK, 2010). The increasing cooperation between MOE and MYCS is showing that ECE is gaining attention and support from the public education system and will be regulated by policies that are regulating public education to some level.

Evidence-based standards and policy-based standards should be on the same page and point to the same values, beliefs, and practices. In most of the cases this alignment exists. However, it is often seen that policy-based standards lag behind what have been found in research or even evidence-based standards. It is important to note that it is not only that the standards themselves are important, but the alignment between standards from different perspectives, and the alignment between practice and standards are more critical in ensuring successful implementation of ECE quality rating and improvement. Quality rating should be based on both evidence-based and policy-based standards. The reliability and validity of the rating tools is the most essential component of quality rating and improvement.

Quality Measurement and Rating

The United States. The measurement that is used to assess quality often has implications for the definition of quality in the U.S. The Environment Rating Scales (ERS; Early Childhood Environment Rating Scale Revised; Harms et al., 2005; Infant/Toddler Environment Rating Scale Revised; Harms et al., 2003 School Age Environment Rating Scale, Harms, 2013, and the Family Child Care Environment Rating Scale, Harms et al., 2007) are the primary measurement tools used to assess program quality. The ERS family of measures are thought of as measuring the global quality of classrooms or homes. They generally are thought of as incorporating two dimensions: structural and process quality. Structural quality describes qualities that are relatively static and stable, such as classroom arrangement, furnishing, and materials. The process dimension describes the dynamic quality such as teacher-child interaction, peer

interaction, and teacher instruction. Since the current study focuses primarily on center-based programs for children under six, when ERS is mentioned it refers only to ECERS-R and ITERS-R. Also because that less research has been done with ITERS-R than that for ECERS-R, most times ECERS-R is discussed.

ERS is used in Statewide Star Rated/Tiered License systems in North Carolina, Tennessee, and Oklahoma, and other quality evaluation and improvement programs in California, Massachusetts, Montana, Mississippi, Kansas, Oregon, Kentucky, New Mexico, Georgia, Florida, Wisconsin, and Nebraska, as well as in Washington, D.C. and U.S. Military service sponsored care facilities (Clifford et al., 2010). ERS has made great contributions in determining environment quality. The scores as well as the report on ESR is the guidance of quality improvement, even though the information acquired from ERS is not specific and sufficient enough for in-depth improvement (Zaslow, Martinez-Beck, Tout, & Halle, 2011).

The psychometric properties of the ECERS-R have been examined in many studies (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005; Perlman, Zellman, & Le, 2004). Most of the studies find that ECERS-R is composed of two factors, each factor roughly composed of items from either a structural or process dimension (Cassidy et al., 2005; Sakai, Whitebook, Wishard, & Howes, 2003). Even though there are variations in factors across different studies, there is consistency that one of the factors is composed of structural indicators, and the other is composed of process indicators. However, the scoring of ECERS-R is neither based on factors, nor on subscales. It is the total score of ECERS-R that is used to represent environment quality in QRIS. As indicated by the

authors, the subscales are not to be used alone, since the subscales do not seem to represent a construct by themselves (Harms et al., 2005). Thus, it makes sense not to use the subscales to represent quality. Rather, subscales are used to guide environment improvement. The accuracy of quality measuring is also largely dependent on the way of using the scores and making sense of the score (Zellman & Fiene, 2012). The internal consistency of ECERS-R is .92 as reported by the authors (Clifford, Reszka, & Rossbach, 2010).

China. As aforementioned, each province or city has their own measurement developed based on *Regulation* and *Guidelines*. The evaluation content includes structural quality, such as physical conditions, teacher-child ratio, teacher education, administrative policies; structural quality, such as healthy practice, nutrition, curriculum, and learning experiences. Take Guizhou, a southwest province as an example; in the total 995 points for the overall quality, the total possible points for administration is 200, for health practice is 200, and for the combination of curriculum and activity is 213. There are also 120 points for child outcome measures, which includes health, social, language, and science/cognitive development. At last, communication with parents counts for up to 82 points (Guizhou Demonstration Kindergarten Selection Standards). It is obvious that structural quality is weighted heavier than process quality.

Little is known about the reliability or validity of measurement tools used in China. Most of the measures that were developed and adjusted according to *Regulation* or *Guidelines* were not validated in large populations. Lately, early childhood researchers have started to create standard measures and tried to validate them in large populations or

through expert consensus (Liu & Pan, 2008; Pan et al., 2010; Li, Hu, Pan, Qin, & Fan, 2014). Pan et al. (2010) tested the validity of a newly developed scale in Beijing. This scale was developed and validated in an earlier study through professional rating (Liu & Pan, 2008). The Kindergarten Environment Rating Scale (KERS) is composed of four domains: (indoor) physical provision, interaction, daily routine and curriculum, including 25 items. In the study (Pan et al., 2010) researchers used KERS in classroom observations to determine the environmental quality, and examined whether the score on KERS differed as a function of programs' ranking and categorizing levels.

Results showed that Demonstration kindergartens scored significantly higher than R2C2 kindergartens and that R1C1 kindergartens had a significantly higher total mean score than R2C2 kindergartens. However, no significant difference was found between the total mean scores of Demonstration and R1C1 kindergartens. Similar findings were found for other subscales of KERS. Thus, this study demonstrated evidence that previously used scales are not valid in differentiating quality, especially for the top levels. It is not clear what the actual measurement is like in Beijing and how it is used in practice. More information may be acquired through this study.

Singapore. Since the ECE quality rating system was just established a few years ago, no data or no academic publication is available showing the reliability or validity of the measurement tool they use in evaluation. Thus, there is a need to investigate if the newly developed tool is valid in reflecting childcare quality.

Quality rating is a critical part of QRIS. The method and the measurement that one selects should match the goal of quality ratings. Quality measurement is the most

flexible and complex part in quality rating. It involves evidence-based items, a psychometrically sound way to produce the score, appropriate combination of different measurement strategies, and high reliability of the assessor as well as the measure. The call for strong data management of the rating outcomes and to start a feedback loop to inform the improvement (Kirby et al., 2014) resonates with implementation science (Fixsen et al., 2005).

Quality Improvement

The United States. Quality improvement is not a one-time event or a product of a series of actions. Quality improvement is an ongoing process that is implemented on a regular basis. It is not as visible as quality assessment, however it is the ultimate purpose of quality assessment—to improve program quality to meet children’s and families’ needs. Valid and effective quality rating should help programs identify what improvement is needed, and should provide programs with accurate and rich information on what resources, assistance, or change is needed in order to enhance program quality. There are different types of approaches that are available to support programs making progress. When talking about technical assistance people refer to a variety of activities that aim for supporting and facilitating program improvement, such as on-site coaching, mentoring, workshops, and professional development opportunities. One study (Neuman & Cunningham, 2009) found that even one-on-one coaching or mentoring is more effective relative to other approaches, the combination of both coaching and other forms of technical assistance is the most effective approach in enhancing teachers’ disposition in valuing children and child development, and professional skills in classroom. Other

than the help and assistance from outside, the leadership of the program director are equally, if not more, essential to program quality.

Research suggests that leadership is crucial in affecting program quality (Fuller et al., 2003; Lower & Cassidy, 2007). In the study by Fuller et al. (2003), questions were asked about whether the community features, such as the income level in the neighborhood, local resource supply, or the director's ability to find resources might have an influence on program quality. The findings suggested that center quality—at least for structural indicators like teacher-child ratio—was not dependent upon the economic strength or poverty levels of local neighborhoods. Higher levels of public subsidies were associated with modestly higher levels of quality, and centers that reported more activities with the local Resource and Referral agency displayed lower group sizes, which indicated that director's ability to look for support did matter. Lower and Cassidy (2007) also found positive relationship between program climate and program quality. Program climate is largely related to leadership. These results revealed that leadership as well as infrastructural support was important in sustaining high program quality.

Whitebook et al. (2012) systematically looked at who was working in early care and education infrastructure organizations. The authors demonstrated some descriptive background information of staff working for QRIS. One of the most important findings showed that overall the employed staff in the infrastructure organization had high levels of formal education. However, more than half of them lacked any ECE-related education and experience. This finding caused concerns in the ECE field. Since leadership is crucial for the success of programs and the whole system, it is important to make sure that

people in these roles know about children and early childhood education. The authors suggested that, as a start, standards of competency for infrastructure staff are needed. The basic standard should be the requirement of “address[ing] child development and pedagogy for teaching young children, ECE systems, adult learning, organizational development, and advocacy” (Whitebook et al., 2012).

A direct way of knowing about the effectiveness of QRIS is by assessing the implementation or existence of QRIS in states and comparing the difference in program quality across time. The Quality Improvement System (QIS) in Florida Palm Beach County employed a systematic strategy which includes a package with quality assessment, professional development, financial support and parent education pieces. This study examined quality improvement in childcare programs through years of implementation of QIS in this county (Ma et al., 2011). The findings indicated that centers entering QIS later were associated with larger improvement in space and furnishing, personal care routines, language reasoning, and activities. The later a center entered QIS the faster the increase in their scores on ECERS-R. The QIS was launched in 2002 in Palm Beach county, and the study was done simultaneously as it was launched. Thus the later programs entered the system, the more mature and complete the system became. This study proved that QRIS is effective in encouraging and helping quality improvement, at least at the structural level (e.g., physical space, personnel).

China. The improvement of quality is not directly linked to the rating of quality. In some provinces programs are encouraged to do a self-evaluation before the outsider evaluation. Based on the results of self-evaluation, the program will be given several

months to make improvements. Other than this, the improvement is implemented by scattered instructional activities that are delivered by local DOE or other private agencies. According to an interview with 880 teachers or administrators from several provinces, 16.2% of the programs received no instructional support in the past year, only 55.4% of the programs were satisfied with the support, and only 37.1% thought the support was effective (Liu, 2012). These results indicate that the purpose of evaluation and ranking is not to inform improvement of quality, but rather for a management purpose, which is to rule programs under a certain system (Liu, 2011). Linking to the call for more evidence-based quality measurements, improvement of quality needs equal if not more fundamental research studies that could inform effective quality improvement in ECE programs. An infrastructure that can systematically support ongoing improvement is essential.

Singapore. Singapore developed a strong program to promote teachers' and directors' professional development. An effort toward developing professional pathways for teachers includes the Fundamental Course for childcare assistants, the Intermediate Course for childcare teachers and the Advanced Course for center supervisors, which was recently changed to a Diploma in Early Childhood Care and Education for teachers and another Diploma in leadership for program directors (Choo, 2010). In recent years, the major focus of policy was on professional development of the work force in ECE (Ang, 2012). In 1999, the MCYS set up the Childcare Qualifications Accreditation Committee, which was later combined with MOE to form the Preschool Qualification Accreditation Committee (PQAC; Choo, 2010). Its role is to assess and accredit early childhood teacher

training courses in Singapore up to the Diploma level. It is unique that the PQAC not only focuses on teacher qualification, but also includes qualification of the trainers of teachers. It ascertains that the courses conducted by teacher training agencies meet the prescribed standards (PQAC, 2008). However less is known how does the current quality rating system (i.e., SPARK) relate to the professional development in Singapore.

Based on the literature there is apparently room for the Chinese QRIS to form a more efficient infrastructure to support teacher and director professional development. The gap may be what is preventing the formation of QRIS under a Chinese cultural and educational background. Questions for the U.S. may be how to be more effective in making use of the feedback loop to inform improvement, and what is the expected qualifications and composition of staff in the infrastructure. Since less data were found in Singapore's case, there is more to be discovered about the effectiveness of each initiative in helping teachers and directors develop. Besides technical assistance, financial incentives for programs and for individuals are equally, if not a more important piece in supporting potential progress.

Financial Incentives

The United States. Financial support to an ECE program in the U.S. usually comes from three resources: Consumer tuition (57%), private sector (4%), and government funding (39%, Stoney, 2014). Consumer tuition is the major resource for most programs in the U.S. However, full-enrollment is not seen in many programs. Program directors may struggle to collect tuitions and fees and also to attract investment from the private sector, which has a very limited share in the total financial resources that

are required. Despite that there are many types of initiatives to distribute financial support, these financial resources are scattered and have different goals and strategies (Stoney, 2014). QRIS should aim to help align the resources to maximize their use. For example, the Child Care Licensing, Pre-K Monitoring, NAEYC Accreditation may use QRIS as the basis for monitoring for applicable standards, while private foundation grants may use QRIS as criteria in their grant making, and State Dependent Care Tax Credit may offer higher credit for choosing higher quality programs (Mitchell & Mathias, 2014). Some of the approaches mentioned above have been implemented in some states already.

Financial incentives that are available for programs nowadays are quality improvement grants, quality achievement awards, tiered subsidy reimbursement, and wage and retention award for programs, and scholarships for individuals (Mitchell, 2012). Research on the effectiveness of financial incentives is limited. Within a national accreditation system, Gormley and Lucas (2002) found that differential reimbursement of 15% made a significant difference in helping programs meet accreditation standards, while a 10% reimbursement made no difference. Within QRIS, however, there were only two studies located that looked at the impact of financial incentives. The most recent study (Yazejian & Iruka, 2014) used a longitudinal design to look at program quality over 5 years in Miami-Dade County, Florida. They found a significant increase in high level quality (i.e., four- and five-star) programs, and a decline of low-level quality (i.e., one- to three-star) programs. The total points programs earned was positively associated with the duration since they entered the QRIS. When looking at financial incentives on its own,

there was an interaction between financial incentives and the duration. The relationship between financial support and total points was stronger for programs that were newer in the system. A descriptive study that conducted interviews with 48 program directors in 9 states revealed that financial incentive was not sufficient in achieving and maintaining high quality (Schulman, Matthews, Blank, & Ewen, 2012). The relationship between financial support and program quality is ambiguous at this point. The current study will probe the mechanism of financial support in North Carolina more specifically.

In China, the financial support for ECE also comes from three resources: Tuition, private investment, and federal investment. Federal financial investment to ECE in China is very limited. The particular investment to ECE only accounted for less than 1.5uo% in the overall investment to education by the year of 2013 (Luo, 2013). Full-enrollment is commonly seen in the majority of programs (Liu & Shi, 2011). However, resources are not distributed equally across programs. More than 30% of the total investments to ECE are distributed to the top-ranked programs, which are called “demonstration programs.”

Take the Southeastern province, Guizhou, as an example. Financial incentives to ECE programs include incentives for quality improvement and subsidies for programs that provide service to low-income families and their children. However, financial incentive is not associated with quality rating. The criterion for “improvement” is based on different criteria than quality, such as if the financial situation of the program is improved (based on personal communication with a program director in China). Less is known in regard to the details of the criteria and the process of financial reimbursement.

In Singapore, financial incentives include governmental subsidy to assist low-income families to pay for childcare programs. This support, however, does not connect to program quality. Since SPARK was only set up in 2010, financial incentives associated with childcare quality may be seen overtime.

Financial incentives are not directly linked to program quality in China and in Singapore as far as the literature shows. In both China and the U.S. program directors are facing the problem of seeking financial support which comes from multiple resources. The alignment of rating and financial incentives is critical to support the improvement process. More information is needed from each country to learn the path of financial resources to childcare programs.

Consumer Education

The last component of QRIS is consumer education. QRIS is designed to inform parents' choice of childcare programs, and to use resources to assist parents in making wise choices in the range of their ability. No research has found on North Carolina, China, or Singapore investigating specifically parents' awareness of QRIS. However some related studies may have implication for consumer education of QRIS in each country.

The United States. Research from the U.S. shows that families select child care for a variety of reasons, including proximity to home or work, length of care needed, and—particularly for low-income families—affordability (Schulman et al., 2012). The view of quality may also vary among parents. For example, a study showed that parents are less likely to use QRIS as a guidance in their choice of childcare centers (Tout et al.,

2010). Compared to referring to the data that shows the quality of teaching, curriculum, or teacher-child interaction, many parents prefer trusting what they see as a warm and nurturing caregiver (Tout et al., 2010).

Less is known about how much parents are aware of QRIS in general, or rely on QRIS to help them find and be able to afford a high quality program. Schulman et al. (2012) in their report provided some descriptive evidence of parents' perception and use of QRIS. By interviewing 48 ECE program directors, the interviews revealed that most parents have limited knowledge about QRIS, and rely more on word of mouth and referral from family and friends rather than searching for official and professional assistance. Some directors felt that using financial subsidies as a way to advocate could be an effective way to attract parents' attention to QRIS. This, in turn, may help parents to notice and start to learn more about what quality looks like in ECE and have their own judgment in choosing high quality programs for their children. Advocating for financial subsidies that provide families with important assistance in paying for high quality programs may cause parents' to pay attention to quality rating instead of relying on other's recommendation (Schulman et al. 2012). According to the study, many parents are not aware of either the existence of the financial subsidies or their eligibility for financial subsidies.

China. Studies have been done to probe parents' perspectives on the definition of childcare or parents' satisfaction toward the childcare program their child enter. In China, research (Lu, 2014) showed that most parents believe that a good teacher is the key factor in determining the quality. A good teacher should have following characteristics: Caring,

respecting children, having good educational strategies, being a good model for children, and spending plenty of time on children's individual development. Parents also thought that a good program should be guided by well-developed pedagogical philosophy that better matches what the parents' value for their children's development. If the program is "famous" according to their relatives and friends was also an important criterion in choosing programs. It is also interesting to note that when interviewing some grandparents, the criteria for a good program was not quite the same as what parents' reported. The largest difference was that grandparents valued the safe and healthy environment over other characteristics when evaluating a program. This difference, although this was from a small sample, implies a change in parents' perspectives and values toward child development in China.

Singapore. In 2012 in Singapore, 1395 parents took an online survey about their view of preschool. The findings showed that most parents (73%) were satisfied or very satisfied with the preschool in which they enrolled their children. However, the same number of parents (73%) were not satisfied with the accessibility, affordability, and quality of early childhood education in Singapore overall. This disparity between the two indices showed that parents are careful in choosing early care and education settings for their children, but at the same time they are very much aware of the development of ECE in Singapore. One of the parents said that it would be better if there was a system set up with defined standards for quality and the fees corresponding to that quality level, because they had no idea what other programs were like and if they paid for what their children deserved. It seems that parents in each of the countries need more professional

guidance in their choice of early care and education programs for their children. What is more important, however even less realized, is what resources parents can get from government or initiatives to enable their choice of higher-quality programs.

Integrating QRIS Components under Implementation Science

Each of the five core components of QRIS, namely, quality standards, quality rating, quality improvement, financial incentives, and consumer education, needs to be aligned with and supported by other components in order to build a strong system that would move ECE quality forward. Based on the literature in each country, each component may be seen here and there, but none of the systems are fully developed. In North Carolina, quality standards and ratings are well developed and established. Quality improvement and financial incentives are supported by outreach initiatives (e.g., CCR&R) as well as within system initiatives (e.g., data system). The alignment between standards and ratings and improvement, including financial support is linked and appears to be effective to some degree. Advocacy in public, especially for parents is one of the gaps that appears to be needed to be filled. Although many of these components appear to be in place, the functioning of the overall system and gaps between the components and the drivers requires further investigation, particularly to inform next steps in the improvement of the QRIS.

In China, quality standards are switching to more evidence-based and child-centered perspectives. Although the quality rating was developed many years ago, it is not yet evidence-based, and the reliability and validity of rating tools and criteria are largely missing. Quality improvement, financial incentives, and consumer education is

either missing or very vague. Much more effort in improving each component, and especially in aligning components with each other appear to be needed. This study will provide further information on the current status of the system in one region and may provide guidance on next steps.

In Singapore, quality standards and ratings were established based on child development, although little is known about the rating tool's reliability and validity in large populations due to its newness. While professional development is strongly established, it appears to have little connection with quality rating at this point. Financial support does not appear to be associated with quality rating explicitly. Consumer education may also need to be strengthened especially after the rating system runs for a while. Overall, each component has promising development, yet, needs more connection between each other to build a strong system.

Implementation science offers a framework to guide the alignment among the components. Quality standards, which are based on empirical research, provide the guiding principles for quality rating and improvement, and need to be aligned with overall ECE policies in each country, state, and district. Quality improvement should be based on the standards and quality rating, and needs to be supported by a wide range of initiatives with resources. The data system, which is driven by the ratings each time, should provide solid evidence for quality improvement. The actualization of improvement, either at the program level or at the individual classroom level, needs to be supported with sufficient and appropriate resources that are associated with the demographic of children enrolled as well as the quality rating. An infrastructure is

essential to make sure the training for staff who are going to carry out the rating and technical assistance is solid. At last, the leadership in either the program or in infrastructure is the key to successful implementation.

Based on the literature and the theory, there is a lack of knowledge in the details of components of QRIS, especially in questions that link to implementation drivers, such as staff selection, training, data management, and leadership. The current study aims for having a more comprehensive understanding of the implementation components and processes of QRIS in each country/region. The study will also take into account the cultural and historical backgrounds of the regions to determine the implementation stage of each QRIS.

Research Questions

Based on the literature review and the guidance of implementation science, the following research questions will be specifically explored in this qualitative study.

Research Question 1: What are the components of each region's QRIS?

Research Question 2: What are the processes of the implementation of QRIS?

Research Question 3: How does the implementation of QRIS affect teachers and families in ECE programs in each region and how satisfied are providers, teachers, and parents with this process?

Research Question 4: What implementation stage is QRIS at in each region, and why?

CHAPTER IV

METHODOLOGY

Selection of Methodology

Qualitative method is the umbrella term for a group of methodologies, such as case study, ethnography, narrative, phenomenology, and grounded theory (Creswell, 2013). All of the methodologies mentioned share some core features, such as having little conditional control over the context, coming from natural observation, usually generating deep understanding of the investigated subject, and having limited power in generalizing to other populations (Padgett, 2012). Qualitative method assumes dynamic reality, which means the reality is not by itself, it is always influenced by the context, and only intensive engagement can make sense of it (Padgett, 2012). Thus, qualitative method has an open system that allows questions to be asked in an open-ended manner and lets things happen naturally in its embedded context. The aim of the present study is to understand the components and mechanisms of QRIS that are currently operationalized in each country/region. The purpose is not to generalize the results to other countries/regions. Thus, a narrowed yet intensive investigation of the details of the operation and implementation is needed. The fact that the QRIS is a complex unit that interweaves all kinds of contexts and dynamic interactions makes it an open-system, which a qualitative method is most suitable to investigate. A qualitative method is ideal for this type of study compared to a close-system when contextual conditions are controlled.

The philosophical assumptions lying behind qualitative methods include an *ontological* assumption, which describes the nature of the reality, an *epistemological* assumption, which describes the relationship between researcher and that being researched, an *axiological* assumption, which incorporates the interpretation and the researcher's value into research, and a *methodological* assumption, which describes the process, and language of research (Creswell, 2013). These assumptions are used to guide qualitative research. Each type of qualitative approach should have a set of assumptions based on the paradigm each type of method falls into.

The selection of the specific approach (e.g., case study, ethnography, narrative, phenomenology, and grounded theory) should reflect research questions and concept framework (Creswell, 2013; Maxwell, 2013). There are several key criteria in determining if a study should be approached with a case study method: (a) it is viewed as a contemporary event; (b) there are clear boundaries for a single case, such as geographic and time boundaries; (c) requires a full variety of evidence (Creswell, 2013; Yin, 2009); and (d) is geared toward answering "how" and "why" questions (Yin, 2009). The major research question of this study is to understand the implementation of QRIS based on a preexisting theoretical framework, and also to use the empirical evidences to complete the framework under multiple contexts. First of all, QRIS is a contemporary event. It is currently taking place in all three countries that will be under investigation in this study. Second, there are clear boundaries for any single case. Each case is distinguished from others because they are embedded in three different contexts. Third, to accomplish the tasks the researcher needs to collect multiple resources, ranging from surveys, to

interviews, and to document review. Finally, questions will focus on how the systems are operated, and why they may lead to different outcomes. The current study meets the four criteria of a case study. In line with the goals and nature of the research questions, the most appropriate approach to answer the questions of this study is a case study approach.

Procedure

Data collection for the current study involved three countries/regions. For convenience sake, the order of data collection was: Beijing, Singapore, and North Carolina. The primary researcher of this study traveled to Beijing and Singapore to collect data. For the administrator interview, the researcher met with the person face to face for 1 to 1.5 hours. Surveys were distributed and collected after interviews took place. The interviews for the program director and teachers were scheduled on the same day. The interview with directors were face-to-face lasted from 30 to 60 minutes. The interviews with the teachers were either completed as a focus group meeting in which both teachers contributed to a conversation regarding their values, feelings, and knowledge of QRIS or as an individual interview when only one teacher was available. Teacher interviews lasted from 20 to 40 minutes. All the interviews were conducted in an indoor environment where participants could talk without disturbing others or being disturbed.

In every interview, basic information was shared first to make participants feel comfortable with talking to the researcher. Once a conversation was started between participant and the researcher, questions were asked (Maxwell, 2013). All the interviews were recorded using a built-in recorder in a phone or tablet with consent from the

participants. Notes regarding the process of the interview were taken during interview procedure.

Before meeting with the administrator and program directors, the researcher asked for any electronic documents, brochures, and websites, etc. that might be relevant to the region's QRIS. Upon meeting with them, the researcher asked for any written materials that might help to explain their QRIS.

Sampling

Interview Sampling

As recommended by qualitative researchers (Creswell, 2013; Maxwell, 2013) the number of participants to interview should range from 3 to 15 individuals. In the current study, 7 to 8 individuals were interviewed in each country/region. Specifically, in North Carolina, one administrator who supervises QRIS and one implementation leader in the state were interviewed. Both in Beijing and Singapore only the administrator who supervises QRIS was interviewed. In both Beijing's and Singapore's cases, the QRIS administrator and the implementation leader were the same person, thus, only one person at this level was interviewed. Throughout the whole document the term "administrator" refers exclusively to leaders in government or agencies, while "director" refers exclusively to program providers. Each administrator received \$50 for her participation.

In North Carolina, at the program level, two program directors from two early childhood education centers were interviewed. In North Carolina, 4 or 5 stars are considered as high quality, and 1 to 3 stars are considered as low quality (North Carolina QRIS). The primary researcher sent out emails to invite programs in the local area to

participate. Eventually a five-star program and a four-star program accepted the invitation to participate. In Beijing, programs in the 1st rank (out of 4), and in the 1st category (out of 3), or who are categorized as demonstration program are considered high quality programs. Programs that are in or below the 2nd ranking, and 2nd category are thought of as medium to low quality programs. The administrator of Early Childhood Education department of MOE chose three 1R1C programs to participate. They were also at demonstration level which is beyond 1R1C rating. In Singapore, the two programs were also chosen by the administrator of early childhood education in MOE to participate. In Singapore, one of the programs was a childcare program and the other one was a kindergarten (see Table 1).

Table 1

Interview Program Information

	Program	Interviewed Director (N)	Interviewed Teacher (N)	Program Size		
				Classrooms (N)	Children Served (N)	Staffs (N)
NC	A	1	2	12	186	43
	B	1	1	8	199	-
Singapore	C	1	2	5	-	-
	D	1	2	12	400	-
China	E	1	1	12	350	69
	F	1	2	8	240	-
	G	1	0	20	671	108

At the program level, three teachers were recruited from the two programs in North Carolina to be interviewed. Three teachers were recruited from two of the three programs in Beijing to be interviewed. Two teachers were recruited from each of the programs in Singapore. All the teachers at least had one direct experience with quality rating. Program director received \$40 and teachers received \$30 for participating in the interview.

Survey Sampling

Efforts were made to gather a larger number of surveys from teachers in each place to get a broader perspective on the teachers' understanding of their QRIS. Surveys were delivered within each program that participated either through Qualtrics or through a paper-based questionnaire. In North Carolina one of the programs participated using Qualtrics and the other used a paper-based questionnaire. Similarly, in Singapore, one of the programs used Qualtrics and the other used a paper-based questionnaire. In Beijing, all three programs received the paper-based questionnaire. In each program one of the individuals who completed the survey won a \$20 gift card from a lottery drawing at the end of the data collection.

In North Carolina (see Table 2) among the 16 teachers most of them were lead teachers ($n = 15$). Mean age of the participants was 35 years ($M = 35.18$), among which 27.8% had less than 2-year degree, 44.4% had a 2-year degree, 16.7% had a 4-year college degree in either early childhood education or child development, and 11.1% had a 4-year college degree in another major. Ten teachers self-identified as African-American, 6 as White, and 1 self-identified ethnicity was Hispanic.

Table 2

Teacher Demographics (Survey)

	Teacher (N)	Mean Age (y)	Race & Ethnicity	Education level			
				<2-year degree	2-year degree	4-year degree (Other major)	4-year degree (ECE or CD)
NC	16	35.18	White: 6 A-A: 10 Hispanic: 1	5 (27.8%)	8 (44.4%)	3 (16.7%)	2 (11.1%)
Singapore	13	36.38	Chinese: 3 Malay: 1 Indian: 5	6 (46.2%)	1 (7.6%)	0	6 (46.2%)
Beijing	42	30.39	N/A	0	12 (28.6%)	29 (69.0%)	13 (31%)

In Singapore (Table 2) among the 11 teachers, most of them were lead teachers ($n = 10$). The mean age of the participants was 36 years of age. Five of them identified themselves as Indian, three as Chinese, and 1 as Malay. Six of them had less than a 2-year degree, one of them had a two-year college degree, and another six had four-year degrees in early childhood education or in child development.

In Beijing, as recorded in Table 2, among the 42 teachers most of them were lead teachers ($n = 40$). The mean age of the participants was 30-year-old ($M = 30.39$). Twelve of them had two-year degrees, 13 had a four-year college degree in either early childhood education or child development, and 29 of them had a 4-year college degree in another major.

Position Statement

As aforementioned, the primary researcher is an outsider of QRIS. However as a researcher of QRIS some ideas are formed and rooted in the knowledge and belief system of the researcher. There are two threats to the validity of interviews as noted by qualitative researchers that must be attended to in studies (Maxwell, 2013; Padgett, 2012). The first is reactivity, which refers to the potential effect of the presence of research on participants' reaction to questions. The second type is researcher bias. For example, it is possible that the researcher might feel upset when hearing people say QRIS is not important, which may result in a change in the researcher's question asking and ultimately leads to an invalid or biased result. Another potential bias lies in the comparison of QRIS in different countries. Since the researcher has more knowledge about QRIS in North Carolina than in the other two countries, there is a tendency to think

that QRIS should look like it is in North Carolina. It is easy to judge QRIS in other countries without taking cultural context into consideration.

To address the first threat, it is important to build a trusting relationship between the researcher and participants (Maxwell, 2013). In the introduction phase of the interview, the researcher was careful to describe the purpose of the study, the use of the results, and participants' rights and confidentiality. However, as argued by Maxwell (2013), trying to minimize the influence of one's presence is not a meaningful goal to pursue. What is more important is to understand how the researcher may affect participants, and include it in the discussion of the results. With regard to the second threat, researcher bias must also be addressed. I acknowledged my bias throughout the study. However, I also conscientiously remained open to participants' experience and feelings about QRIS in all three countries and avoided asking leading questions.

Another way to minimize researcher bias is through triangulation (Maxwell, 2013; Padgett, 2012), which is to collect data in multiple ways. In the current study, a portrait of QRIS will come into being from three resources: the QRIS administrators, ECE program directors, and ECE teachers. Both the interview and surveys were used with directors and teachers respectively. The overlap between the two formats allowed for a type of validity check of the responses and views. Finally, to minimize bias during interpretation a second researcher read all the transcripts from the interviews and verified the derived themes.

Measurement

The measurements that were used in the current study included document review, surveys and interviews. Both the survey and interview protocol were developed jointly by the primary researcher and her advisor based on the research questions, and were guided by the theoretical framework, which was implementation science.

Document Review

The documents were requested and collected from the QRIS administrator(s) and program directors. Documents included an assessor training manual from North Carolina, the DCDEE website, a measurement scale from Singapore, and two online documents about rating standards from Beijing's kindergarten.

Survey

The survey was only completed by program teachers. The survey asked about their knowledge, feelings, and beliefs toward QRIS (see Appendix 1). There were four sections in the survey. The first section was about their general knowledge and feelings about QRIS. An example question was "To what degree would you say you are familiar with your state/district/country's Quality Rating and Improvement System?" All the questions in the first section were on a 1 to 5 Likert-scale from the least to the most. There were 5 questions in the first section. Questions in the second section were about their knowledge and feelings specifically to quality rating. An example question was "To what degree in general do you think the QRIS result was accurate?" A follow-up question was how accurate they thought specific ratings were, like the environment rating, curriculum rating, and program philosophy rating. Most of the questions in the second

section were on a 1 to 5 Likert-scale from the least to the most. The third section asked about their knowledge and feelings specifically toward quality improvement. An example question was “To what degree would you say you are familiar with quality improvement?” and “In general, how would you rate the effectiveness of the improvement activities?” All the questions in the third section were on a 1 to 5 Likert-scale from the least to the most. The last section asked about their satisfaction toward QRIS from multiple aspects. Example questions were “How satisfied in general were you with the standards used in the assessment?” and “How satisfied in general were you with the support you received to improve program quality?” The last question asked participants to rank the effectiveness of 4 commonly seen improvement activities. The 4 activities were: professional development, one-on-one coaching, financial support, and workshops. There were total of 18 questions in the survey. The survey is located in the appendix.

Interview Protocol

The interview protocol for administrators and for implementation leaders was the same set of questions in order to have triangulation on data. The questions involved two major sections (see Appendix 2). The first set of questions asked about the components of QRIS in their state/city/country. The second set of questions was related to the process of implementing the QRIS. The questions were structured based on the three drivers of implementation science and the definition of the implementation stages (Fixsen et al., 2005). In the protocol, questions were organized under three implementation drivers, namely: Competence driver, organization driver, and leadership driver. Most questions

were to be answered with facts that the participants knew of. An example question was “What policies/standards are used to align QRIS?” However, as is true in many cases, the “fact” is what is perceived by the participant as the “fact.” It represented their values and perceptions on the matter. There were a few questions that explicitly went beyond asking about facts, and involved more personal perspectives and attitudes. For example, a question was asked as “From your perspective, how do the *Rating* and *Improvement* go hand-in-hand?”

The interview protocol for directors and teachers overlapped greatly with the survey. The purpose in conducting an interview in addition to survey was to acquire richer information behind their beliefs and feelings toward the QRIS. Using different methods to collect the same information also increases the validity of the data.

The initial questions in these interviews were very broad and general (see Appendix A). As the conversation went along more probes were used to facilitate the conversation (Creswell, 2013; Maxwell, 2013). For example, as a follow-up question to participants’ previous descriptions of the big picture of QIRS, the researcher asked specific questions regarding quality rating standards and measures used in the rating. The order of the questions generally followed the same protocol in each region. Since there was overlap in the questions, it was normal for participants to talk about something that was to be asked later. It was important for the researcher to be very familiar with the protocol in order to ask follow-up questions without following the order on protocol (Creswell, 2013). The interview questions are located in Appendix A.

All interviews were transcribed by the researcher or an individual from an independent transcription company. The interviews and surveys from North Carolina and Singapore were written and conducted in English. The interview questions and surveys for Beijing were first translated into Chinese by the lead researcher, who is a native speaker. All the interviews were conducted in Chinese and then transcribed into written Chinese text before being translated into English.

Reliability and Validity

Since only the primary researcher coded and analyzed the data, the inter-coder agreement (Creswell, 2013) was not achievable. However, a second researcher read the transcripts, helped review the coding, and examined the themes created during data analysis. Validity was achieved through triangulation among multiple forms of data as mentioned in position statement.

Data Analysis Plan

Data from the document review and administrator interviews were drawn upon to answer the first two questions regarding the general status and history of QRIS in each state/district/country, and the implementation process of the QRIS. Data from program directors and teacher interviews and teacher surveys were used to answer research question 3 which asked how the QRIS affected teachers and families and how satisfied they were with the process. The fourth and the fifth research questions were answered based primarily on the administrator interviews, however, all the information that was acquired in this study was made use of for these analyses.

The first step of data analysis was to listen to the interview records, read the interview notes, and documents (Creswell, 2013). After transcripts were created from the audio recordings, the researcher read the transcripts for several times to get a sense of the whole picture (Creswell, 2013). NVivo 10 (<http://download.qsrinternational.com/Document/NVivo10/NVivo10-Getting-Started-Guide.pdf>) was used in coding. The coding in NVivo is very straightforward. To answer questions about how QRIS was implemented in each region, I first drew themes from the implementation framework. Themes were also referred to nodes in NVivo. They will show on the left panel. To code, I selected and dragged sentences or paragraphs from the imported document, which is shown on the right panel, to the nodes. The implementation science framework provided the basic themes for coding, such as staff selection, staff training, and facilitative administration. However themes that were associated specifically with a certain country or region were created as well. Since only a few teachers and directors were interviewed in each region I did not ask NVivo to count the frequency and generate themes. All the themes from director and teacher interviews were created by me based on my perception of the frequency and depth of their feelings.

Pattern matching is one of the most preferable analytic strategies in a case study approach (Yin, 2009) and in the current study the outcomes were guided by the theoretical framework. I coded information that was related to the framework and made connection among them. The categorizing strategy, which is usually referred as coding and classification, involves comparing and contrasting, and looking for significant statements from pages of transcriptions of the interview (Creswell, 2013). The first step

was to identify significant statements that could be coded under each variable and link them to other variables. However, it was equally important to keep the answers open-ended. Since the questions were exploratory there could be some themes that were missing from the existing framework. The second strategy was to code the important information from all the resources, even if they were not relevant to the framework that the study was based on. The themes should not be formulated as abstract concepts, but rather as condensed description in order to disclose the meaning (Lindseth & Norberg, 2004).

The purpose of the third research question was to match the independent variables with the dependent variable. In this case, the independent variables were derived from the first two questions. The dependent variable was the current status of the whole system. To be specific, the first two questions answered how the “machine,” the quality rating and improvement system, worked, and the third question sought to find out if this working system was efficient in producing the desired outcomes. In order to strengthen the internal validity of this study, alternative explanations for the dependent variables were identified through reviewing documents. However the purpose was to make a strong argument that none of the alternative explanations could systematically and comprehensively explain the dependent variable (Yin, 2009). If this match fails then it means the current theoretical framework is not the best in explaining what happened, it either needs to be adjusted or there may be other theories that could better explain the outcomes.

CHAPTER V

RESULTS

The results are presented in the order of research questions for each country separately.

North Carolina

Components of QRIS in North Carolina

The Quality Rating and Improvement System (QRIS) in the U.S. is composed of five components, which are standards, measurement, improvement, financial incentives, and customer education (Tout et al., 2010). However, as could be imagined, the current components have been through an evolving process since the system was established.

When the original QRS was launched in 1999, there were three components in the standards, including compliance history, teacher education, and program standards. In 2006, compliance history was dropped as an independent component since it was not considered as a valid indicator to distinguish high from low quality. It is now included in program standards (admin-1; admin-2). A 75% compliance history is one of the prerequisites for further quality evaluation. Any program that is below 75% will not be eligible for star rating. This change resulted in a two-component standard instead of a three-component standard system. Thus, the current system includes program standards and education standards, which are tallied as quality points. Within the program standards are compliance history, ratio and group size, and environment quality. The three criteria

measure the basic and structural quality of a program, and are worth up to 7 points in the rating. Staff qualifications are also worth up to 7 points. The other bonus quality point came from either curriculum, family partnership, or administration and management.

One of the most recent changes in the standards was that curriculum is no longer counted as a criterion or quality point. This change took place in 2014. Previously, a curriculum list specified the qualified curricula that programs could use to earn the quality point. However, as the number of qualified curriculum increase the child care commission felt many curricula could be included on the list. So instead of adding to the list the committee decided that it is more effective to examine the curriculum used in each program to see if it is aligned with the North Carolina Foundations for Early Learning and Development (Foundations), as well as Developmentally Appropriate Practice (DAP). One administrator stated, “it went from 14 approved to almost 40 approved. It doesn’t mean as much as it used to” (admin-1). It is not that curriculum will not be looked at during the assessment, rather it is that using an appropriate curriculum will not add an extra point to the assessment score.

Besides the two major changes, the standards have remained relatively stable across the 16 years of implementation. Some problems, however, did emerge over time. With the current system some programs may choose to not have an observed assessment of quality because they have enough points from the education standards to reach a higher star level (admin-1). Even though this is an unusual case, and it is not possible for a program to achieve 5-stars without being assessed, but it did happen with some lower-level programs (admin-1). There are programs that are not participating in a quality

assessment at all beyond receiving the compliance license because they have enough children enrolled due to their reputations among parents. Another identified problem reported by administrator-1 was that programs may not receive credit for some efforts they make, for example, cultural competence practice, and inclusion strategies. Neither of these practices are specifically included in the evaluation standards. The presence of these problems may lead to more severe problems in the long run. Future directions will be discussed later.

The second component of QRIS is measurement. Measurement is composed of two dimensions: one is the measurement tool; the other is the rating structure. When the system was launched in North Carolina in 1999 the Environment Rating Scales (ERS) were selected as the environment assessment tools to be used in homes and centers. The star rating structure is based on points in which the final score is the sum of the points from the two-component standards. If the results from the Environment Rating Scale observations are high enough, then the scores can be used to add points toward the star rating in the program standards component. There has not been change in the measurement over the years. Some evidence shows that ERS may have limited power in differentiating quality (Hestenes et al., 2014). A new tool is under development by a group of early childhood education professionals from three different universities in North Carolina, Delaware, and Kentucky. This project is funded by the Race to the Top Early Learning Challenge and aims to measure program quality in a more comprehensive and evidence-based way (document review). There is also a trend toward including more assessment tools to form a toolkit (admin-1). More evidence would be needed no matter

if the new tool, or the toolkit, or the combination is going to be used in future assessments. However, evidences have shown that the current version of ERS may not sufficiently differentiate environment quality (Hestenes et al., 2014; Zaslow et al., 2011).

The “I” in QRIS refers to improvement. Improvement does not only happen after an assessment is done but may also occur beforehand. There are consultants from the Division of Child Development and Early Education (DCDEE) who help with the basic quality maintenance, and technical assistance (TA) through coaching or mentoring. The most commonly mentioned problem by administrator-2 was that there were not sufficient resources to cover all the programs that made requests for TA support (admin-2). As a result, many 4 or 5 star programs are not able to receive TA support at times when there are requests from 1 or 2 star programs. “We focus a lot of resources on programs that have never been assessed or programs that [have] lower star[s] and that’s where the bulk of the TA money has gone” (admin-2). In another example this administrator gave, one TA provider spent a year in a program that received extremely low score on ITERS-R in their last assessment. Considering the limited resources this is a significantly long period of time. In general, the use of resources always prioritizes weaker programs. The philosophy behind the decisions is to move lower programs up so that no child ends up in a really low quality program. However, the problems it creates are obvious too. “There’s a concern that that reduces motivation in some ways for higher performing programs. It’s like once you do well then you’re on your own. There’s not a lot of support for you” (admin-2).

The same situation applies to consultants. The duty of a consultant is similar to that of TA providers. They provide professional advice for programs to maintain their quality and also to make improvements. They are not like TA providers, though, since they go to every program regularly rather than only upon request. So every licensed program will have a consultant come in at least twice for one scheduled and one unscheduled visit. However the consultants need to visit 7800 programs across the state. Their time for each program is “amazingly limited” (admin-2). The lack of resource in providing technical support as well as consultancy could be a barrier for more effective improvement at this stage of development.

Compared to the minimum technical support that high-rated programs receive, the high-level programs receive more financial reward than the lower-rated programs. Back in 1999 all programs could receive some financial support no matter what the quality level they achieved. Since 2008, however, only programs with a three-star rating or above could receive financial incentives in the form of a subsidy. The higher the program rating the more financial support per child with a disability or child in a low income family that program would receive. Needless to say, this mechanism did work in encouraging programs to strive for higher star levels. This brings two benefits. First, programs will have more enrollment, which means more income, because children at risk (i.e., with a certain disability or from low income families) are more likely to enroll since they receive more financial support by enrolling in higher-level programs. The second benefit is that it encourages families who have children with disabilities to enroll their children in high quality programs. For example, if the commute is not as convenient as a low quality

program close to their homes, the financial benefit that enrolling in a high quality program brings may exceed that consideration about commuting. Problems associated with the current incentive system is, as aforementioned, some 1 or 2 star programs prefer to stay where they are because they have enough enrollment already (amdin-1 and admin-2). Another problem, though not as salient as the former one, is that some programs are simply driven by the financial incentives to increase the star rating. “A lot of folks are doing these things because they have to get subsidy money, or to get a higher star. We see a lot of this external pressure. Teachers being told, if you don’t score a five, you could get fired” (admin-2). Contrary to technical support that prioritizes lower-quality programs, financial support is viewed as a reward for high quality programs.

The fifth component, customer education, maybe the least emphasized in North Carolina among all five, and is also viewed as one of the “biggest challenge[s]” (admin-1). It is challenging because it is hard for the public to see the value of high quality early childhood education. What is more basic is to educate parents as well as the public about developmentally appropriate practices. Parents who lack that knowledge may not be aware of the difference between a high quality and a low quality program. These fundamental changes may take a long time and need not only the effort from QRIS, but the whole society. However, parents sometimes are just not aware of their choice about programs. Efforts were made to better inform parents in North Carolina about their choices. For example, “On our website, we have ‘choosing appropriate childcare’ and it gives you questions (for parents to ask when entering a childcare)” (admin-1). Besides these efforts, there are still many parents who exclusively rely on “appearance of the

building,” “emotion reaction to the setting,” or “word of mouth” (admin-2). A good sign is that program directors noticed that more and more parents were starting to ask about the quality rating and mentioned some indicators of it (admin-1, admin-2). However it is noteworthy that customer education is viewed as a component that parallels the other four, indicating that it has not been given enough attention, partly due to the difficulty of implementation and the lower investment in this component.

The most urgent problem, however, may be that the current system is losing the power to distinguish high quality and very high quality programs. After decades of implementation, 85% programs are now three-stars or above. High quality programs (4 and 5 stars) need a “roadmap” (admin-2) to guide them to a new height. Due to the increasing demand for high quality childcare and accumulating research evidence, a revision to the QRIS is taking place. According to one administrator, it is usually normal for the QRIS to be revised after 10 to 15 years. In 2009, ten years after NC QRIS was launched, DCDEE convened a group of stakeholders and experts across state to think about the future of QRIS in North Carolina (DCDEE). An advisory committee got together to look at the research evidence associated with QRIS as well as what was currently happening in the state. After 3 years’ effort in 2012 a summary report came out with 300 recommendations for improving the whole system made by the adversary committee (admin-1; DCDEE). Before implementing those changes DCDEE is making phone calls to stakeholders asking them about their feelings and perspective in making changes based on the recommendations.

The validation study took those 300 recommendations and looked at, ‘Well, what could we do with our current facilities to see what might make a change in quality?’ That’s what we’re doing right now is we’re doing phone interviews with a hundred facilities, 75 centers and 25 homes, and asking them questions about professional development plans. Do you have them? If you don’t, how hard would it be for you to implement them? Would that be something you would like to get points for on a star rated license? Just for an example. (admin-1)

The implementation of the recommended changes may take years to complete (admin-1):

[W]e’ll have to prepare lots of things. We’ll have to prepare a FISCAL note, which talks about what the financial impact would be: Positive or negative. We’ll have to prepare the research and the literature saying why that’s best practice . . . Then there’ll be a transition time for facilities to come into compliance.

As discussed in this section, there are identified problems associated with each component. However, the five components of QRIS are not operating independently. The implementation of the system is the key to coordinate the five components and to produce desired outcomes.

QRIS Implementation in North Carolina

Star rating process. One star is the mandatory star level in North Carolina. In fact, when a program opens it needs to prove that it can be operated legally and meet the health and safety requirements. When the program is licensed it also automatically receives a one-star rating. The one star was temporary in the first six months. In the six months programs that have “enough in program and education to raise their stars” have the time to prepare and apply for the assessment (admin-1; admin-2). If programs stay at one star quality they will be officially a one-star program by that time. If they think they

can be higher and manage to demonstrate a higher quality they will get a raise in their star level. Thus, all the programs in North Carolina are one star or above.

Other than the licensing phase, star rating is composed of two parts. First the consultants will look at all the criteria in the standards other than the environment, and second the assessors will be assessing the environment using the ERS. The total score will be calculated based on the two components. Any program that wants to be assessed will need to apply for it through DCDEE. The consultants who work with the program will forward this application to the environment assessment agency. The program will then be given a month's window, during which the assessors can show up on any day to do the ERS assessment (admin1; admin-2). On the assessment day, each classroom is randomly picked based on the number of classrooms in each age group. One third of the classrooms are assessed, and if there is only one classroom for a certain age group that classroom will be assessed (amin-1). Other than doing the on-site observation, there is an interview either before or after the observation.

Programs also receive at least once announced visit, and one or up to two unannounced visits per year from DCDEE for the other components in the rating standard, (admin-1). Thus the structural part, such as teacher qualification, family partnership, and administration and management are assessed on regular basis. After the ERS assessment, a report will be sent to the consultants who work with the program. The consultant will either directly forward the report or review the report with program leader and teachers (admin-2; director-1).

Quality improvement process. The quality improvement process is not an obvious or even necessary process for some programs. But generally if a program has questions or problems the director will find support from consultants or technical assistants. A consultant typically helps with looking at if the program has met the basic standards and helps maintain quality (director-2). The support that consultants provide are usually available around the time of programs' licensing renewal. Technical assistants help with more specific problems and could go in more depth in all ways, such as environment setting, teacher-child interactions, or administration. TA only come to help upon request by programs unless a noticeable issue was raised, such as an extremely low score in ERS.

Implementation science demonstrates a framework of how a program or system may integrate different parts to generate desired outcomes. It is composed of three dimensions, which include competency drivers, organization drivers, and leadership. Generally speaking, the organization driver provides the big picture; the competency driver is the agent that drives the system; and leadership is monitoring and problem solving. QRIS is a complex system that involves multiple agencies and different levels of support and effort. In this section the working mechanism of this system will be split into three pieces for the convenience of discussion. However it is important to keep in mind that they are interdependent parts and in reality should be viewed in conjunction with one another.

Competency driver.

Staff selection. People are the agent of a system. They are the ones that drive the machine. Generally speaking, the staff in the NC QRIS are assessors, consultants, and technical assistance (TA) providers. Specific selection criteria apply to each position. Assessors are required to have at least a bachelor's degree and working experience in an ECE program. Both education and experiences are emphasized. One administrator explained that having experience in a classroom will actually "make a difference in [the quality of the] report [writing]" (admin-2). However, she also pointed out that too many experiences could be harmful as well, since that may create bias for the assessor when she walks in a classroom to do an assessment. There are different levels or certifications associated with assessor positions. Anchors, reliability checkers, and supervisors are the higher level positions. "Supervisors supervise the assessor, the person. Reliability checkers supervise the quality of work that the assessor does" (admin-2). Previously the supervisor position required applicants to have management working experiences. Later on people noticed that it was important for the supervisors to have assessment experience as well. Thus in the present selection criteria the supervisors are selected from the pool of assessors (admin-2). Reliability checkers are also previous assessors. They are promoted to be the checker because they average high reliability, have "good communication skills, really good knowledge of child development, [and are] able to provide development rationales for the indicator requirements" (admin-2).

Selection criteria for the consultants and TA providers are similar to assessors. Consultants and TA providers are required to have at least an associate's degree. TAs

have expertise in different domains. For example, some TAs are specialists in curriculum, some are in environment, and some are in promoting teacher-child interactions. There are also generalists who can have an eye on the big picture.

Staff training. For the assessors, staff training, official assessments and performance assessments are woven into an ongoing process that can be tracked all the time. A 190-page staff procedure manual lays out the detailed training procedures. Basically training starts on the first day when a newly hired potential assessor starts to work. It is a series of observation trainings, including all the age groups, and all the program types (e.g., centers, homes, inclusion-programs, non-inclusive programs, and one- to five-star programs). It usually takes 2 months or more to complete the training. By saying complete it means the reliability reaches a minimum criterion (85%) with a reliability checker (admin-2; manual). After the assessor is “signed off” to do official assessments they will still do regular reliability checks with the designated reliability checker. There are different ways to decide when the check is needed. It could be based on frequency or time, depending on the reliability status that assessor achieved each time (admin-2; manual). They view “each reliability check as a training” (admin-2). Thus training is woven into daily work, and performance assessment is woven into training.

Training for consultants is not as systematic as it is for assessors. It involves a mentoring and coaching based training strategy. During the training period the new consultant will go out with a supervisor who will check on the consultant’s work. “We do what’s called basic job skills training. A lot around the rules and how to go out and look at the rules. There’s a lot of hands on stuff that takes place” (admin-1). Consultants, like

ECE teachers, also acquire training through workshops, webinars, and information sharing. TAs are thought of as a part in the QRIS system but they are not directly hired or trained under DCDEE. There are different agencies that provide TA and they are trained differently. There is no formal communication tunnel among assessors, consultants and TAs. This sometimes creates problem that will be discussed later.

Organization driver.

Facilitative administration. Facilitative administration refers to a series of practices that prepare the environment for a system's installation and development. Metz et al. (2013) recommended the following practices to be included in a well-functioning facilitative administration: (a) Ensuring leadership is committed to the new program or practice and available to address challenges and create solutions, (b) developing clear communication protocols and feedback loops, (c) adjust and developing policies and procedures to support new ways of work, and (d) reducing administrative barriers to ensure a hospitable environment for high-fidelity program (or practice) implementation.

QRIS in North Carolina is right at the point when an update to the system is needed (admin-1). That is why an advisory committee was formed to review QRIS in North Carolina and work on giving practical recommendations to DCDEE. For the first part, ensuring the commitment of leadership to the system, the leading institution, DCDEE, was in place for years before the QRS was launched in North Carolina. There were several changes that happen during the early years. For example, in 1999 a star rating system was launched to replace the A and AA rating system. In 2005 the compliance history was dropped from the standards. It is obvious that the leadership was

working in pushing the system forward. From individual level, both administrators I interviewed had great experience in early childhood education. For the first administrator even though she just started working in her position 18 months earlier, she had 20 years of experience working in ECE field. For the second administrator, she has been working in the same agency for 16 years, since the agency was established in 1999. They were both highly qualified people in decision making and problem solving.

For the second aspect, developing clear communication and feedback loops, evidence showed that the communication between advisory committee and DCDEE is effective. For example, the advisory committee, the one that is formed to review the QRIS, needs DCDEE's assistance to search for evidence to generate the recommendations. DCDEE then utilizes the recommendations to make improvements. From the program's perspective, the communication is initiated by DCDEE to ask about their feelings toward the change. However, the communication between the assessment agency and improvement agencies are not as effective. In fact, this happens to be a long existing problem within the system (admin-2). Rating and improvement are both essential components of QRIS. However evidence showed that they do not necessarily go hand-in-hand in practice. The communication between rating and improvement are usually indirect through programs. The interpretation of the assessment report mostly relies on the program itself, consultants, or the TA. The fact that the three parties (i.e., assessors, consultants, and TA's) are not trained under the same system makes the communication even harder (admin-2). Programs may receive different instructions from the consultant

and TA on many issues (admin-2). A feedback loop among the three parties is necessary to make sure there is successful implementation of rating and improvement.

There is a greater need for policy adjustment to support the new way of work. However even before the turning point, adjustments need to be in place. For example, Montessori programs are currently not included in the rating and improvement system in North Carolina because of some of the conflicting philosophies between the typical programs and Montessori. Efforts are being made to include these kinds of programs into the QRIS. But rather than trying to change Montessori programs, practitioners and experts need to devote their energy to finding ways to change the ERS to fit with Montessori programs (admin-1). The other example is that agencies that are receiving different funding and at different levels should share research and funding with each other to better support programs in that region (admin-2). In order for a systematic update to take place much more change and adjustment needs to happen. Administrator 1 talked about the necessary change in fiscal plans, and looking for empirical evidence to support the new system: “we’ll have to prepare lots of things. We’ll have to prepare a fiscal note, which talks about what the financial impact would be” (admin-1).

Finally, reducing administrative barriers and creating a hospitable environment for implementation should apply to every domain of QRIS. Technology enables data collection and utilization in the system. The assessment agency uses a website that incorporates every detail of the assessment, and tracks every training and official assessment. The data from all the activities can be retrieved and used easily. Both administrators expressed the feeling that the lack of staff working as consultants or TAs

is one of the biggest barriers in implementing the “T” successfully. This remains an issue for now. Facilitative administration should make sure that funds are distributed to the appropriate places, there are clear policies and protocols to follow, and the communication loop is established and well-functioning. It is not possible for a system to be perfectly implemented at once. But with a prepared and supportive environment, it paves the ground for greater success.

System intervention. As one of the components of the organization driver, it is parallel to facilitative administration. Facilitative administration focuses on building a supportive system environment, while system intervention concentrates on supporting this attempt by aligning and communicating with external agencies and making use of external recourses. It incorporates the following practices: (a) forming and supporting an early childhood leadership team, (c) developing formal processes to ensure policy-practice feedback loops and bidirectional communication across and within levels of a system, and (c) engaging and nurturing champions and opinion leaders (Metz et al., 2013). It is actually hard to define what is “external” and what is not. Since QRIS is not a single program, it is a complicated system that involves layers of agencies. However it is initiated and organized by DCDEE, thus agencies such as R&R, and Smart Start may be considered as external agencies and resources. Other less ambiguous agencies and resources are federal level policies and funding, media, and societal groups.

Looking for funding and support is an ongoing process for any non-profit system or agency. It is not an easy task for ECE though. Due to a lack of immediate effect and repay for what has been invested, more money is invested to solve immediate problems

rather than preventing problems from happening. One of the administrators mentioned that the political environment is now improving since Obama is supporting early childhood education, either from a policy level or funding level (admin-1). Although it is still a tough mission it is important to advocate the value of ECE to the public and establish communication loops between practitioners of QRIS and policy makers. The cooperation between the advisory committee and DCDEE is a successful example of such loop.

Inform decisions—Decision-making data system. Data, including fidelity data, short-term, and long-term data, are collected to support the assessment process. In North Carolina reliability data are recorded each time and can be easily tracked online by supervisors and assessors themselves. These data directly inform the reliability check strategy in the short-term. In the long-run it also helps improve the whole training protocol. For example, administrator-2 shared a change that had been made on a reliability check procedure:

Yeah, we learn and learn and learn and learn. Even initially, like we were checking people every six or 10 times forever like even if we changed it in 2006 but still at that point, we might have had people who'd been doing assessments for seven years and they were still being checked every three weeks. Then we looked at it and we're like, "Oh, reliability doesn't vary. This is extra work." That's when we developed the two frequency based and time based . . . (admin-2)

The utilization of the long-term data actually helped save time and energy for both assessors and the whole agency. This type of change is enabled by a hospitable and supportive environment that is set by facilitative administration.

There is not a systematic data pool for the whole system yet. However according to admin-1 one of the Race to the Top projects is working on establishing a system-wide data plan. It is supposed to include both assessment and improvement data. With time the feedback loop between assessment and improvement may be strengthened and used to inform better decisions.

Leadership.

Adaptive leadership. Adaptive leadership is observed when competing perspectives or philosophical problems emerge within a system. In NC QRIS's case, adaptive leadership happened a few times in its history. One of them was changing from AA system to Star Rating system in 1999. Another one was to drop compliance history as a rating standard. The up-coming revision is another case of the demonstration of adaptive leadership. The solution to those fundamental problems usually requires experimental trials. The current step that the QRIS committee is taking is to pilot each recommendation by interviewing stakeholders to collect their opinions involving motivation as well as resources for change to have an understanding of the big picture. Other than that, validation studies with restricted criteria will be conducted to test more some key changes, such as a new measurement tool or new approaches. Based on the investigation a lot of things need to be prepared: "In doing so, we'll have to prepare lots of things. We'll have to prepare a FISCAL note, which talks about what the financial impact would be. Positive or negative. We'll have to prepare the research and the literature saying why that's best practice" (admin-1).

Adaptive leadership may involve a collection of leaders rather than a single leadership. To be efficient, different divisions or institutions need to take different responsibilities. However this is to be determined along the process of updating the system. “The commission may choose to take one piece, the whole package, half the package, who knows” (admin-1). At last, there will be an integration and alignment of different components.

Problem solving—Technical leadership.

Technical leadership. Technical leadership is observed when low level of disagreement emerges and problem solving does not involve a change of structure or values. This kind of leadership should be happening on a daily basis. Any adaptive leadership should involve a variety of technical leadership. An example of technical leadership in QRIS is that the assessment group gradually learned from their experiences and made changes on training protocols and schedules. Technical leadership enables a system’s daily operation.

Integration and compensation. It is important to keep in mind that the drivers cannot function independently. The components within each driver are also interdependent. The organization driver sets up the environment for implementation. However it is not possible without the support from competency driver and leadership. The same principle applies to the other two drivers. Within the organization driver, for example, facilitative administration will not function well without the input from the data system or the support from system intervention. On the contrary if the facilitative

administration does not provide a hospitable environment for implementation, the alignment with external resources may not be successful.

However, components can sometimes compensate for each other. For example, when the selection of staff is not implemented as planned, staff training may compensate it by conducting trainings based on the features of this group of employees. Thus, the success of implementation is determined jointly by all the factors involved.

In general QRIS in North Carolina possesses strong ability in its utilization of leadership. The leadership role maybe taken by different groups while there is one overarching leadership group that oversees the whole system. The most developed driver is the competency driver which makes sure that this system is driven by a group of qualified and reliable people. With the strong support from both leadership and highly qualified staff this system also possesses ability in making necessary accommodations and adjustments to better align with a new system. The data system provides empirical support for decision making and changes. One of the pitfalls of this system is the loose connection between R and I. Despite the frequently used database in supporting decision making, the fact that only short-term data is often used makes it less strong.

The Effect of QRIS Implementation in North Carolina

QRIS in North Carolina was launched 16 years ago. As described previously it is a heavily evidence-based system that collected and utilized data to inform decisions. From a macro-level of view, after years of implementation most of the programs, including family care and other types of programs, are licensed and rated and also supported by the system in different ways. By July 2014 North Carolina had 7140

regulated childcare facilities, serving 249,654 children, of which 83,700 benefited from subsidy support (DCDEE website). To promote wise as well as convenient choices, the DCDEE allows individuals to easily find a childcare program with the specified star level and/or the location of different program options. Research is consistently conducted to better inform policies and practices.

From a micro-level of view, program directors and teachers are the ones that are having direct contact with the system. They are the targets of assessment and also the builders of quality in a program. Directors' and teachers' feelings and experiences have important implication for the system's growth. To what extent they think the system is effective may shape the direction of QRIS development in the future. In the following section results from both surveys and interviews with teachers and directors will be presented.

Survey Results.

Descriptives. The first question asked about their knowledge of a series of components of North Carolina's QRIS. As Table 3 illustrates, the teachers reported a moderate to high level of overall knowledge about the QRIS (3.94/5). Among all the components of QRIS, most teachers thought that they were very familiar with the standards on teacher-child ratios (4.53/5), staff qualifications (4.05/5), and professional development associated with improvement (3.95/5). Teachers were less confident in their knowledge about financial incentives (3.21/5).

Table 3

Descriptive Survey Results

		North Carolina (<i>n</i> = 16)	Singapore (<i>n</i> = 13)	Beijing (<i>n</i> = 42)
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Teacher Knowledge	Overall	3.94 (.75)	3.86 (.77)	3.25 (.80)
	Staff-child ratio	4.53 (.61)	4.00 (.68)	3.09 (1.01)
	Staff qualification	4.05 (.78)	3.93 (.62)	3.39 (.90)
	Self-appraisal (Exclude NC)	-	3.36 (.84)	3.44 (1.08)
	Group evaluation (Beijing Only)	-	-	3.09 (1.03)
	Outsider assessment	3.84 (.84)	3.29 (.83)	3.06 (1.11)
	Feedback	3.78 (1.00)	3.29 (.73)	3.65 (.91)
	Professional development	3.95 (.78)	3.64 (.93)	3.89 (.82)
	Financial incentives	3.21 (.79)	3.00 (.88)	3.22 (.97)
Teacher Satisfaction	Staff-child ratio	3.53 (1.50)	3.71 (.73)	3.69 (.93)
	Staff qualification	3.68 (1.46)	4.00 (.56)	3.85 (.76)
	Self-appraisal (Exclude NC)	-	3.75 (.62)	3.83 (.88)
	Group evaluation (Beijing Only)	-	-	3.78 (.83)
	Outsider assessment	2.68 (1.00)	3.75 (.75)	3.86 (.90)
	Feedback	2.89 (.99)	3.92 (.64)	3.87 (.78)
	Professional development	3.42 (.96)	3.77 (.83)	3.96 (.78)
	Financial incentives	3.05 (1.4)	3.42 (.79)	3.57 (1.04)

Table 3

(Cont.)

		North Carolina (<i>n</i> = 16)	Singapore (<i>n</i> = 13)	Beijing (<i>n</i> = 42)
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Parent Knowledge	Program rating level	3.37 (.96)	2.93 (.92)	3.92 (1.03)
	Staff-child ratio	3.16 (1.26)	2.71 (.99)	2.96 (1.03)
	Staff qualification	2.68 (1.06)	2.64 (.84)	3.15 (1.07)
	Self-appraisal	-	2.22 (.44)	-
	Outsider assessment	2.53 (.96)	2.36 (.75)	2.85 (1.07)
	Self-appraisal	2.47 (1.21)	2.64 (1.08)	3.04 (1.10)
	Outsider assessment	2.26 (.99)	2.64 (1.08)	2.65 (1.06)
Teacher Feelings	Affect my work	4.00 (1.25)	3.36 (.93)	3.50 (.91)
	Promote program quality	4.11 (.94)	3.86 (.54)	4.06 (.83)
	Stressful for me	3.42 (1.47)	3.00 (.88)	3.36 (.86)

In the second part of the survey teachers answered questions regarding their satisfaction with the QRIS. Overall this part of survey received lower scores compared to the questions about knowledge. The most satisfied component was the staff qualification standard (3.68/5), followed by teacher-child ratio (3.53/5) and professional development (3.42/5). The least satisfied component was the assessment process (2.68/5).

Teachers were also asked to answer questions about parents' knowledge about QRIS from their perspectives. This part of survey received low to moderate scores overall. The highest score of this domain was parents' knowledge about the rating level of the program (3.37/5), followed by teacher-child ratio (3.16/5). The lowest score of this domain was parents' knowledge about financial incentives (2.26/5) and about how professional development relates to quality improvement (2.47/5).

Teachers were asked to rate their general feelings toward QRIS as well. In the first question: To what degree is their daily work affected by QRIS, the average score was 4 out of 5. Teachers rated the second question higher: to what degree they think the QRIS helped the program improve (4.11/5). According to the third question, teachers were moderately stressed by QRIS (3.42/5).

Teachers rated the accuracy of their program rating quite high (4.42/5). The overall effectiveness of professional development in improving program quality was rated 4.06 out of 5, which falls between "effective" and "very effective." When asked about each type of professional development opportunity, 16 teachers had workshop or meetings experience, 14 had experience in sharing resources with peers, colleagues or other programs, nine had continuous education experience, and nine teachers had one-to-

one coaching experience. Continuous education experience received the highest effectiveness score (4.5/5), followed by workshops and training (4.41/5), despite its low frequency. One-on-one coaching experience received the same score as the score for sharing resources with others (4.38/5).

Correlations. The correlation (Table 4) between teacher perceived knowledge about staff qualification and teacher perceived pressure in assessment was significantly negative ($r = -.60, p = .006$), indicating that teachers who had more knowledge about staff qualification requirements were less likely to be stressed by the assessment. Similarly, teacher's satisfaction toward teacher-child ratio was negatively related to teacher perceived pressure ($r = .48, p = .036$), implying that the teachers who were satisfied with teacher-child ratio were less likely to be stressed by the assessment.

Overall teachers reported that they were familiar with the QRIS standards and process. They were less satisfied with some of the components, such as the assessment process and feedback. However, they did not feel a lot of pressure when being assessed and they did think that the system could help enhance program quality. In general, teachers thought professional development was effective, mostly with continuous education and workshops. Teachers who had more knowledge in staff qualification or teachers who were more satisfied with teacher-child ratios were less likely to feel stressed during assessment.

Table 4

Correlations between Teacher Satisfaction and Teacher's Feeling about QRIS (Singapore)

	Teacher Satisfaction About ...						
	Teacher-child Ratio	Staff qualification	Self-appraisal process	Outsider assessment process	Feedback	Professional opportunities	Financial incentives
Affect your work	-.18	-.60 [*]	-.60 [*]	-.45	-.22	-.61 [*]	-.36
Enhancing program quality in your program	-.11	.26	.63 [*]	.52	.20	.28	.36
Pressure create for you	-.12	-.48	-.77 ^{**}	-.64 [*]	-.29	-.67 [*]	-.44

 $p < .05^*$, $p < .001^{**}$

Director experience. Two directors from two programs were interviewed in North Carolina. Director A was from a five-star program, and director B was from a four-star program. Both programs are for-profit programs. Program A has 12 classrooms, providing both full-day and half-day care for 186 children, and have 43 staff for now. Program B has 8 classrooms that serves 199 three to six years old children. Both directors have worked in the current facility for more than 10 years, so they both had a great amount of quality rating experience. When talking about the QRIS the two directors had quite different reactions. The predominant themes from director's interview are twofold: One is confidence and gratefulness, and the other is concerns and challenges. Special attention was paid to the matching between feelings and previously described QRIS components as well as QRIS implementation.

Confidence and gratefulness. Director A was very confident about her program. She felt little pressure if an assessor was coming in to assess the program. This confidence came from her perspective of the available resources. "I think we're lucky in this program that we have that kind of access to people outside the classroom. Lots of programs don't have that" (director A). It seems that she had no financial stress, so that it was relatively easy for her to improve the structural environment and provide more support for teachers' professional development. Team building was another reason behind her confidence: "We have team leaders for our infants and toddlers, we have a team leader for our twos and we have a team leader for our pre-K." During the course of talking director A mentioned several times "I feel lucky," and expressed concern for "other" programs that do not have the luxury of resources. It implies that she was very

much aware of the difficulty of maintaining or improving a program's quality, that is why she felt lucky because she knew if not for those resources the program would have a hard time scoring well in the assessment process.

Both director A and director B were very grateful for the fact that the program had managed to hire a person to supervise or monitor the program quality all the time. The person, who is in a consultant or TA's role, is hired full-time by the program. They were referred to as supervisor or director by director-A and -B respectively. His/her job is to help with quality maintenance and improvement by examining each room, observing teacher-child interactions, or occasionally doing an assessment based on the scales. Director A felt that she did not even need support from the DCDEE consultant or TA provider: "I don't know that we necessarily need technical assistance for somebody to come out and help us, but I think that's because we've got people in place that are able to serve [in] that [role] for us." Director B indicated that the supervisor provided a second set of eyes to monitor the program, and the existence of such a role actually took some of the burden off her shoulders. The little use of TA services matched on to the previously described trend that high star level programs are having less access to TA due to the lack of TA resources. However even though the facts looks the same the reasoning was quite different from outsider's view (i.e., administrator) and insider's view (i.e., directors)

Concerns and challenges. The concern that director B had about raising their star level was mainly due to her teachers' education levels. This appears to be the most challenging aspect for her program to achieve: "That (staff education) is truthfully the only challenge that we have. That if I could get 75% of my staff with degrees, then that

would be great.” She realized that it was hard to hire a highly educated person while they have better paying options in the public school system. The financial situation in her program would not allow many teachers to get more education or training. Although the teacher attrition rate is not a concern for director B, because she felt that teachers in this programs were paid higher than their counterparts who had the same education level, environment setting also costs money, and this was a big issue for director B: “It’s more of buying materials, and we don’t have a whole lot of money to do that.” This resonates with what director-A was worried about and how “other” programs who do not have as many resources could make improvements.

There was an inconsistency in director-B’s concern. On one hand she indicated that teacher education was the biggest challenge, yet in her interview she actually had many concerns about the measurement as well. The biggest concern that director B had was that teachers were rated on multi-tasking and the way they were rated did not allow any failure in any of the tasks: “we’re all human, and reminding, okay, let’s get into this mindset of telling children what we want them to do, and not telling them not to do something. That’s the hardest thing with my staff.”

Director A demonstrated concerns about the teachers who had never experienced the assessment process before. For her, the most challenging aspect was to get teachers ready for the assessment, and particularly to make sure that they got used to the process so that they were calm during the assessment. She, like director B, also had financial concerns when it came to environment settings. However, for her, it depended on how

much better she wanted the environment to be. In another words, it was not as much a concern for her as it was for director B.

Director-A expressed fewer concerns in terms of what happened in her program, instead, she was more worried about the validity of the current system. “I feel like it (assessment) is easily manipulated” (director A). She felt that three-years was a long time for quality to vary for some programs. In addition, a few hours of observation and a short interview could not represent the full picture of the program quality. She was very concerned that some five-star programs were not actually five-star quality.

Teacher experiences. Two of the teachers (teacher A and teacher B) that were interviewed were from the five-star program (program A). The other teacher (teacher C) was from the four-star program, (program B). Compared to the distinct feelings that each of the director held, the teachers’ perspectives and feelings were more closely aligned.

Nerve-racking. Teachers’ experiences related to the quality assessment were mixed. It is obvious, though, that they all found the assessment challenging, especially on the evaluation day. The word “nerve-racking” was frequently mentioned. The idea that they were being observed and evaluated itself was putting some pressure on teachers even though they were very sure what they were doing was correct:

Oh God, it’s so scary. And it shouldn’t be because we’re doing everything right already and we were a hundred and ten percent prepared, and we knew that we were prepared and we knew that we were doing everything right, but it’s still really nerve-racking to have somebody watch your every move. (teacher A)

Other reasons may also create this tension in teachers on evaluation day. Two of the teachers (teacher B & C) felt that the assessor did not take children’s behavior into

account, but only focused on the teachers' behavior. For example when children did not "behave" in the classroom, teachers felt they scored lower on some of the items. The teachers felt that it was not their fault in making that child behave that way. They are nervous about having a crying baby or fighting peers in the classroom. One of the teachers (teacher B) said she always worries about "are they going to behave that day, or are they all going to throw fits and tantrums, and you worry that the assessor looks at that."

Another reason for being nervous is due to the fast changing schedule in a childcare classroom. Teachers were afraid if they were off schedule they would be "counted off" on the rating scale. Every time when a teacher moved to a new activity, she had to adapt to the pressure of being watched in this new activity. They were always fearful about whether they were missing anything on the scale or if the assessors missed what they had done.

Two of the teachers (teacher A & C) had only been through the assessment once. Their expression of the nerve-racking feeling was stronger than the other teacher (teacher B) who had been through assessment twice in her career. Apparently the more experienced teacher was more confident about what she did. All she worried about was whether children were going to behave on that day. The other two teachers, however, worried about the schedule and, for one teacher, about if she was following the book (scale). It seems that having assessment experience helps teacher to understand the expectations and handle it with more confidence.

Satisfied versus unsatisfied. Besides the feeling of being nervous on the observation day, two of the teachers expressed a high level of satisfaction about the interview after the observation. They felt that the interview was adequate in helping them address issues that were not observed on that day, and helped them to illustrate the rationale behind some of their actions. Two of the teachers were satisfied with the feedback they received: “And they’re usually very specific with the evidence about exactly what they saw or what they didn’t see. So it gives you a good idea of why you got that score” (teacher A).

Other than feeling some standards are not very realistic, teachers did not have anything that they felt really unsatisfied about in terms of the rating. Two of the teachers used the example from routines. One of them expressed annoyance about the hand washing requirements. She thought it was not realistic to require children to wash their hands after any single activity (teacher A). The other teacher found the sanitation requirement too detailed and that made her nervous because she was afraid she missed some steps (teacher B).

Support. The level of support for teachers in each program appeared to differ from one another. Although teachers did not explicitly talk about the support they received, it was implied here and there in their descriptions. One of the teachers in a five-star program was especially confident because she knew she was in a five-star program, so she believed what she was instructed to do must be right: “And I feel like I did okay, we had five stars, so I know I did something right” (teacher A). It conveyed a message of trustworthy and confidence when she talked about her program over the course of

interview. Teacher B was previously working in another program before she came to the current one. She compared the two programs' supportiveness she felt very grateful to be at the current program: "I don't remember being given anything before the assessment . . . Like here, we've been building up for it. They've been talking about this for two or three months and she's not coming until the summer." Both teacher A and B talked about the training they received to prepare for assessment and professional development opportunities that were paid by the program. They all said even if the assessor walked in today they would do just fine because they trusted the supportive system and the training they received.

However, neither teacher A nor B were aware of the external support from the state level. They kind of knew someone would come to do some kind of examination, and they knew when the director had difficulties she would seek help from them. But they both said they did not know who a TA provider was and had never being helped by a TA. This may due to the fact that the program had built a strong supportive system and had hired someone to exclusively monitor the program's quality, so that teachers do not even need the external support.

For teacher C, she found the program to be very supportive in her professional development by giving her time off for school and a small amount of financial support. Teacher C, compared to teacher A and B, was more aware of TA providers. Even though she did not have experience with a TA until then, she heard about others' talk about TA in very positive way:

The teacher who was able to experience that opportunity said it (TA) was very helpful because it wasn't stressful. She wasn't coming in as an assessor, but she did come in with the knowledge of an assessor and how the assessment would be and helped with the classroom to prepare for assessments, so I think that is a good resource when it comes to preparing for assessments. (teacher C)

However she was not satisfied with the supervisor that the director talked about. She felt that the supervisor was not as effective as she was expected to be in promoting the program's quality. Teacher C thought that the supervisor did something that looked really fancy and advanced to people who come to visit, but in terms of actually improving the program quality it was not enough.

One of the purposes of QRIS is to inform and also support programs' quality improvement. Based on the survey it seems that most teachers felt that this system did help improve program quality even though they felt that assessment affected their daily work to a relatively large degree. Interview results resonate with the survey that teachers were nervous most of the time about the assessment. However there is an obvious distinction between the five star program and the four star program in terms of teachers' as well as directors' general feelings about the system. Both teachers and the director in the five-star program conveyed a feeling of confidence when being assessed, while the teacher and the director from the four-star program had more concerns with assessment. According to the interview it seems that the five-star program did not rely on external support to improve, while the four-star program did use some external support in addition to its own built-in support. This difference speaks to what the administrator mentioned that the discrepancy of TA distribution among different quality programs. However according to director A, they chose not to rely on TA rather than having limited access to

TA. It may be possible that it is because this lack of resources that high quality programs (i.e., five-, and four-star) made an decision not to rely on TA, or it could be that they have the resources to hire a TA-like person in their programs to provide regular support.

Implementation Stages of QRIS in North Carolina

Implementation stages represent the degree to which a system has gone on the road to a high fidelity, and comprehensively implemented system. The four stages (i.e., exploration, installation, initial implementation, and full implementation) describes the typical process that a system would take to get to a full implementation. However it is important to keep in mind that there is not a clear division between any two stages, and the direction could also be nonlinear, which means any implementation effort may return to an earlier stage at any time (Metz et al., 2013). It is also important to understand that different drivers may be at different stages at the same time. In the case of the QRIS in North, after 16 years of implementation it has components that develop at different paces.

In full implementation, competency drivers should include effective recruitment and interview protocols, high fidelity in outcome delivery, and training, and effective performance assessment to inform further improvement. In North Carolina QRIS's case, the competency driver is the most developed and stable driver at this point. There is a manual that comprehensively describes the process of staff selection and training. Short-term data are used to inform training and performance assessment as well. The learning and improvement process has become a regular protocol for staff. The competency driver showed great stability and fidelity through the years. The three components of the competency driver (i.e., selection, training, and coaching) are to a large degree

synchronized as well. However, this judgement may only apply to the assessment part since little is known about the selection and training protocol for consultants and technical assistant providers.

As for the organization drivers of QRIS in North Carolina, the picture becomes much more complicated compared to the competency driver. In the full implementation stage organization drivers are supposed to include evidence-based decisions, demonstrate great alignment between practices and policies, have sufficient support from external resources, and strong leadership teams. Given that the revision of the system is taking place there are some reoccurring events that put the system at a mixed stage in terms of its organization drivers. The reoccurrence is taking place in several domains. The formation of the advisory committee is a feature in the installation stage when a leadership team should be forming in the system. However, this advisory committee is not permanent, the enactment of the recommendations will still be done by DCDEE which has been in the leadership role for years. Thus, the formation of advisory committee should be viewed as an effort to support the leadership, and it demonstrates the evidence of a strong leadership focus that makes adaptive changes based on data.

In the process of implementing the recommendations there is an unavoidable return to the earlier stage. For instance phone calls were made to programs asking about their opinions about some potential strategies (admin-1). The new measure is under a development and validation phase. Fiscal plans need to be reexamined and adjusted to meet the changes in the system (admin-1). These steps are usually taken in the exploratory or installation stages. However, this also demonstrates this system's strong

capacity to make adaptations and address administration barriers. Thus facilitative administration is at its full implementation, since it is the ability to address barriers, and make appropriate adaptations that create a hospitable environment for the system.

System intervention describes the requirement for external support and alignment. The biggest challenge that was identified by administrators is the lack of consultants and technical support for programs. They also have concerns about the continuity between assessment and improvement. The fact that assessors, consultants, and TA providers are trained under different systems makes the alignment harder. Up to now there has not been a plan to address either of the challenges or concerns. One of the features of the installation stage is that the external support is loose and lacking alignment. Thus, system intervention could still be at the installations stage.

Short-term data were used to inform decisions in assessment and technical support. However, long-term data are less used systematically so far. According to the administrator-1 a system-wide data system is being built with the support from the Race to the Top funding. In the initial implementation stage data are supposed to support daily practices and administration. In the full implementation stage it is expected to review the data regularly. Thus, the data-based system is at its initial implementation stage and moving toward the full implementation stage.

To view the system as a whole, it seems that the QRIS of North Carolina is at a full implementation stage, since most components are fully implemented, except one at the installation (i.e., system intervention) and one at the initial implementation stage (i.e., decision support data system). It has the fundamental functions to provide a sound

environment for the system to sustain and grow. However with the upcoming updates in the system there has to be another round of implementation process starting from exploration and moving forward. It does not have to start from nothing, however. The sustainability of the existing system shall provide strong support for the new system's development.

Singapore

QRIS Components in Singapore

As aforementioned, QRIS has a relatively short history in Singapore. Therefore the components are basic and clear-cut with less intertwined components. The QRIS in Singapore is composed of standards, measurement, and improvement. Linking back to the five general components of QRIS, there is a clear absence of financial incentives in this system. Parent education is also not considered as a component of QRIS in Singapore <<It is, in the QRS 4.2 B4>>. This may due to political or cultural reasons, it may also due to the early stage that this system is currently in. In the following paragraphs I will present results from the investigation with Singapore early childhood education department administrator (admin-3), ECE program directors, and teachers, as well as from document review.

Standards. The Singapore Preschool Accreditation Framework (SPARK) was launched in 2010, however the development of the standards as well as the measurement started in 2008 (admin-3). The administrator of the early childhood education department in Singapore's MOE was one of the leaders who developed this system. The standards for QRIS are linked to the standards for kindergartens' curriculum framework in Singapore.

In this curriculum standard children are viewed as curious, active, and competent learners (MOE, 2012). The desired outcomes of education are: A confident person, a self-directed learner, an active contributor, and a concerned citizen (MOE, 2012). The standards of QRIS are supposed to support such values which are held by early childhood education practitioners in Singapore.

SPARK standards are illustrated in the introduction of the measurement tool: Quality Rating Scale (QRS). It depicts the framework and components of the value behind the standards. Leadership is the one that drives the other components of program quality. Other components include Planning, Staff Management, Administration, Resources and Curriculum and Pedagogy. The underpinning guiding principles of the framework are: Professionalism with Impact, Innovation with Purpose, and Partnership for Growth that illustrates the professional development for teachers, flexibility and creativity of program, and connection with families respectively. Health, hygiene and safety are included in the regulation, because “it is [a] very important issue, that sets the standard for operational purposes” (admin-3). The desired outcomes is a holistic outcome where “children are excited about learning, interested in learning, and child[ren]’s wellbeing [is emphasized]. Children must learn how to take care of themselves” (admin-3; QRS).

The development of the standards as well as the measurement itself is not directly or merely based on child development or educational research evidences. It is, however, also based on the investigations in places that have a successful model of QRIS implementation or ECE framework: “We started developing the Quality Rating Scale in

2008 when I made visit to countries like the U.S, China, Italy. And my team also had people who visited East Asia, and UK” (admin-3).

In terms of the rating criterion, instead of ranking programs with stars, SPARK came up with a quality-levels system with three levels and an intermediate level in between the first and the second level. The first level is Certification, the next higher quality level is Commendation, and the third level is Accreditation. The determination of the level in which a program quality would fall is not merely dependent on the average score of the rating scale. It also depends on other criterion such as Curriculum Leadership.. For example, when programs first get evaluated and if they reached an average score that corresponds to the first quality level (emerging), the Emerging Level on the QRS they will receive a “Certification,” or be able to say they “achieved standard for SPARK certification.” Since it is the initial implementation of SPARK for evaluation, even if a program’s quality has the potential to be higher they will still be rated on the first quality indicator, and will only receive the first level of rating (i.e., Certification). To promote quality improvement beyond Certification and give recognition instead of ranking among programs (admin-3; director D), SPARK has introduced Commendation in 2015 and all the qualified centres were recognized in retrospect.

Here we don’t encourage ranking of programs, and we don’t think it is suitable to use stars because, um, we think that the star system is seemed to relate to hotel star rating . . . in Singapore we [are] mindful of unhealthy competition ; we will never give them this idea that I am first, you are second . . . (admin-3)

When a program is evaluated for the second time, it is a Re-certification. It is qualified to receive recognition in Progress in Teaching and Learning if the program

shows a significant improvement over the first evaluation. However, for programs that start with higher base in quality, since it is more difficult for them to achieve a significant improvement, they will be qualified for Commendation. The establishment of the recognition in Progress in T & L aims to acknowledge program effort in making improvement and encourage them to make further progress in a few years (admin-3). The highest rating level up will be Accreditation, which will take time to evolve and take place when the quality of the EC sector has arrived. According to the administrator this level will have a really high level of requirement. But since “we are still working on the performing level (i.e., the second level of quality indicator in the QRS) is very difficult really” (admin-3) for programs, it seems that there is still some time to work on the highest level of the pyramid.

In 2011, 87 centers applied, and 69 achieved standard for SPARK at the emerging level, which means they have passed the Certification criterion and were granted certifications. In 2014, the 69 centers were up to the three-year validity to get reassessed, and 59 out of 69 were successfully recertified. The loss of centers is mainly due to the change on the center’s side, such as shutting down or converting from kindergarten to childcare. All centers participated in reassessment have achieved the Re-certification so far.

Measurement. In SPARK the measurement and the standards are more intertwined than separate since most of the standards are measured in the scale. The whole system is set up to measure educational quality in programs, so the target of evaluation includes programs that provide educational learning opportunities;

specifically, only classrooms that serve four- to six-year-old children and bear an educational function. Those classrooms are mostly in kindergartens and some childcare programs with four-year-old children.

The development of the measure made references from a series of published scales. The scales that were used as reference are widely used in the U.S. The administrator also paid several visits to the authors and consulted about reliability and validity issues:

To develop a comprehensive instrument, we looked at ECERS-E, ECERS-R, . . . they are environment rating scales in the classrooms. We also looked at other instruments, like PAS for the organizational, management, staff management, and resources. So we started writing the items in 2008, and we also engage[d] oversea[s] consultants for content validity, one is Dr. Thelma Harms. (admin-3)

Domains in the scale correspond to the standards illustrated in the introduction. They are: leadership, planning and administration, staff management, resources, curriculum, pedagogy, and health, hygiene and safety. The scoring rubric is similar to the Program Administration Scale (PAS; Talan & Bloom, 2004) which has both vertical as well as horizontal scoring dimensions. On the horizontal dimension there are three quality levels along a 6-point scale: emerging, performing, and mastering, each representing a quality level with the mastering as the highest level. On the vertical dimension there are a few indicator strands that are presented in a minimum-to-desired order. The administrator told me that this design will allow program to make progress gradually using this scale as guidance:

So every strand is a progression across and every quality level downwards is set in [a] progressive level, so that every indicator is linked vertically downward too. (admin-3)

Content-wise, also similar to the PAS, the QRS's primary focus is program administration and operation. Items are related to administration, planning, professional development, and curriculum. Even though the framework of the measurement indicates that structure and process are weighed equally in determining the final quality, relatively less attention is paid to process quality, such as teacher-child interaction or teaching practices (QRS; teacher G). However, the purpose of this scale is clearly defined as a measurement of program quality, so it does provide a holistic view of the program.

Improvement. “We always convey to the centers that they don't exist for SPARK. SPARK exists because of ‘you.’ We want ‘you’ to improve. So SPARK is always position[ed] as [an] improvement process, and improvement journey” (admin-3). This is the thing that was mentioned again and again, not only by the administrator, but also by almost everyone that was interviewed. This is the tenet of SPARK. The policy makers do not want to make the system a factory that produces quality centers, but to make it a journey that everyone can join and enjoy. Thus improvement is an inseparable part of the journey.

It is unique that improvement starts even before the first day that SPARK was launched, because it was a totally new thing for everyone and it should not come as a surprise to the EC sector when it was launched. All the program directors and teachers were required to attend a training workshop that was held across Singapore. “In 2010 before the launch we train[ed] up to 4,000 teachers and center leaders on the whole

instrument. It [was] a two-day workshop” (admin-3). SPARK requires programs to do self-appraisal before starting outsider quality evaluation. After the official evaluation was done, programs compared the difference between their self-appraisal and the evaluation done by educational experts.

The report contains information on strengths and areas that need more effort. According to admin-3 the report will not give every detail of the evidence. What is in the report are the descriptions that correspond to the descriptions on the scale with a single example to demonstrate evidence of that score, so that it is clear which domains need improvement and what the goals are, based on the description of each indicator. When the report is ready for a program, the assessors who did the evaluation or other assessors who have high expertise come to the program and review the report with the director and teachers. They answer questions that are raised during the process, and coach them how to improve on each dimension. The QRIS is used throughout all the activities: training, self-appraisal, report review, and improvement.

Along with the implementation of the system, questions and problems emerge, however many are addressed in a timely manner. Program directors talked about a “direct head” that they could call at any time to ask questions (director C; director D). A group of people who work as “quality assurance consultancy” is also available for programs. They are mainly in charge of helping programs get familiar and prepared for the assessment. Typically the consultants will go to programs to assist for five sessions. After the five sessions the program needs to apply for SPARK. The connection between the

administrator and programs is very close. Program leaders may even call the administrator directly when they encounter big challenges (director D).

At this stage of development SPARK has three major components: standards, measurement, and improvement. Financial incentive is not provided. Parents are aware of SPARK through public announcement and through programs. However the purpose of parent education is only limited to the needs associated with parents' cooperation and understanding. At this point, most of the parents are not guided by SPARK when picking kindergartens for their children (director D, teacher F).

QRIS Implementation in Singapore

SPARK rating process. SPARK aims to conduct a quality rating that is beyond licensing, thus it is voluntary for programs to apply. Self-appraisal is required when a program applies for the rating. The official assessment includes document check, interviews, and classroom observations. The document check includes the self-appraisal that was done by the program, the program plan, the curriculum plan, and the action plan (admin-3). Before the assessors visit the program they review all the documents that have been sent to them by the program. Two assessors are in the program to do the observation on a certain day.

Usually a typical timetable will say, morning, they will familiarize [themselves with] the center, by walking around. And then they will go into classroom to see the teaching and [at a] minimum they have to see through a whole lesson, if it is a 45 minutes, they have to see through the 45 minutes. (admin-3)

The two assessors split the work so that they are able to see different things and align the evidence afterward. In the afternoon they interview the program director and

teachers for about 2 hours to “cover strategic leadership, curriculum leadership, and as they move along we want to know what is the role of the center leader” (admin-3). The evidence from the document check, observation, and interview are triangulated with each other. The site visit takes two days to complete. Following the observation, the assessors have a week to write, review, and finalize the report. The reviewing process occurs at three levels. First, a group of assessors review the report, to see if the evidence supports the rating. The report is then to be sent to the “gatekeeper” to make sure that reviewers have gone through everything. Finally, the report is sent to the administrator for a final review and to confirm decisions on any ambiguous scores: “Then they will come to me for moderation on areas that they highlight. I will decide whether these are ‘okay,’ and ‘leave it’ (admin-3).

The report will then be delivered to the program by the assessor who did the assessment together with the quality assurance officer (admin-3).

The launch of SPARK took more than two years of operation and now it is moving toward the fifth year of operation. It provides a unique case study to view how a system is developed, installed, and then initially implemented. In the following section the three drivers of implementation science framework will be demonstrated in more detail.

Competency driver.

Staff selection. “Singapore is a very small nation.” This particular statement was mentioned several times by the administrator. The direct consequence it brings is the lack of manpower to do the work. There are multiple roles that need to be filled from a limited

pool of professionals: assessors, trainers, reviewers, and consultants. Singapore is also a fast-paced country where people work hard and the economy grows fast. According to the administrator they are not afforded the luxury of hiring a new person and spending years to train that person to be qualified to take the job. Thus, the innovative solution was to hire from the Ministry of Education, the retired principals, superintendents and deputy directors to be assessors and reviewers.

The principals, superintendents and deputy directors, however, were not familiar with early childhood education, but they were experts in education and pedagogy, and most of them held a master degree (admin-3). Most importantly they knew about management and leadership, which is considered to be the core driver of program quality: “They are very experienced people, and they have [a] leadership quality also because they were the principals before. So they are familiar with the ‘leadership,’ they have experience in management” (admin-3).

Some of the principals are at their 60s or even 70s. But they “love it” because they get to be reemployed and be useful again. However not all the recruited principals can eventually be an assessor. The performance assessment is done after training to decide who can be certified as assessors. The administrator proudly said that “we are very careful with our selection of [an] assessor, very highly qualified people then can become [an] assessor.”

The quality consultants are different group of people who are from the early childhood education field. They all have ECE degrees, and have great experience in early childhood education as well as management knowledge. It was intentionally decided that

this group of people were not eligible to become an assessor because they are more likely to be biased when doing assessment since they know specific programs and people in this field well. In this high stakes assessment the most important thing is to be reliable and validated. The view is that principals from elementary school have fewer, if any, connections with the kindergarten classrooms, thus they are more likely to be neutral when assessing. “Actually, when we talk about that our group of assessors [doesn’t] know the centers because they are not from the field. So they are very neutral people” (admin-3).

All the assessors and consultants are under part-time employment except for eight permanent assessors. The eight permanent assessors provide the standards in the group. They often provide training to new assessors and also do reliability checks.

Staff training and performance assessment. Training for assessors is very rigorous (admin-3). Since most of the assessors have little knowledge about early childhood education, when they were recruited they started with taking a 100-hour ECE course that was delivered by MOE. Following this, 22 staff members went to the University of Helsinki, Finland to pursue a diploma in ECE studies. By 2014, most of the assessors had a diploma or degree in early childhood education (admin-3). The second part of the training is instrument training. “It takes two days of training of the instrument; two days field practice in a childcare centre and another two days field practice in a kindergarten” (admin-3).

Other than the first reliability check to determine who can be an assessor, each assessor’s performance needs to be reassessed regularly. There was a change from doing

inter-rater reliability check once a year to a standardized workshop three-times a year.

The turning point was when they noticed that they do not have enough manpower to do reliability checks once a year. Limited numbers of assessors need to go to the field to do the assessment, so it is hard to schedule reliability checks. This challenge led to standardization training. The standardization is like a re-training:

The reliability is done by [a] standardization session. That means all the assessors have to come back to attend a workshop, we will go through the instrument, so standardized . . . So it becomes a practice, so three times a year, beginning of the year, before they start work; middle of the year for debriefing, and also for feedback; end of the year for wrapping up. Begin of next year again start. (admin-3)

The training for consultants only involves instrument training since they are already experts in ECE (admin-3). There is no performance assessment for consultants because they are trusted as the expert who knows the field.

Organization driver.

Setting the environment—Facilitative administration. Facilitative administration refers to a series of practices that set the environment for a system's installation and development. In order to ensure the leadership was committed to the new program in Singapore, before SPARK was launched several leadership teams were in place. "The curriculum and pedagogy is developed by the curriculum specialists from the MOE. So we have these experts in curriculum that helped us in there. And for area in health, hygiene and safety, [it] is developed by the regulation team from MOE" (admin-3).

There is also a leadership group that was in charge of connecting all the components together: "We are the quality assurance team—we pull everything together

and we developed 1-4 (constructs), which is leadership, planning, administration, staff management, and resources” (admin-3).

After SPARK was launched different departments were also in charge of different domains. For example, quality assurance (QA) was in charge of assessment and improvement. Other departments were in charge of regulation and professional development. A SPARK Committee meeting is held once a month. People from “all over Singapore,” such as ministry of education and polytechnic, about seven members attend the meeting to look over the report that was written by assessors and raise questions. This process may strengthen the commitment of different leadership groups in SPARK and also make sure the communication is established to allow for feedback.

Adjusting and developing policies and procedures to support a new way of work started during the development of SPARK and will likely continue into the future. When the measurement tool was developed there were a lot of questions and uncertainties around several key issues: What would be the best fit for Singapore’s kindergarten? What is the best way to represent the results? What quality indicators should be used? What scoring strategy may apply? Those questions remained unanswered, not only at the beginning of the system, but also changed along the way during the five years of implementation.

The administrator was very proud when she talked about the development and finalization of the measurement. This process took about two years. As aforementioned, she visited several countries to learn about their scales and scoring strategies to create the new scale that would be used in Singapore. When the new scale was under construction

the developers invited centers to field test the existing items and collect feedback from them.

[The] first time is to see whether our instrument is easy to understand, that one is the first field testing. So we refined the indicators to something that is more comprehensive and um, ease of reading, and also short, meeting people's needs. So we go on to the second field testing. (admin-3)

The second field test was done three months later with the updated indicators. This time the purpose was to test the assessment procedure. "And then we found that 2nd testing is also not desirable because our assessment process did not allow us to see certain things happening" (admin-3).

The first and the second rounds of field testing were with a small sample size. Moving toward the third pilot study, the administrator paid a visit to one of the authors of the Environment Rating Scales and consulted her with content validity issues. The third field test was also the last one before two pilot tests took place, one in 2009 and another one in 2010 before the official launch of SPARK. This time the goal was to test content validity for the scale. Fifty programs were recruited for this pilot study.

An obvious evidence of adjustment and change in policies and strategies could be tracked along the development of the scale. After SPARK was launched, adjustments were still needed. For instance, there was a change in the reliability check policy due to the lack of manpower. The administrator also anticipates changes in the future. The prospective change is mainly due to the accumulated experience they have in operating this system.

When I developed the instrument in 2008, 2009, two years—my observation of the landscape was very low. So when the instrument was developed, the consideration at the emerging level was to focus more on program provision. (admin-3)

In order to create a hospitable environment for high-fidelity implementation it is important to reduce administrative barriers, such as resistance to stakeholders, inconvenience for practitioners as well as for staff. Since SPARK is new to everyone, promoting the buy-in becomes the first issue on the agenda. As mentioned by almost all the directors, teachers as well as the administrator, when SPARK was introduced people had tons of questions for this new system. They were panicked, suspicious, and, mostly, wondered what change it would bring for them. The administrator realized that letting people accept and participate in this system was the first and most important thing for SPARK. So in the measurement development phase, programs were invited to give feedback on the scale. In this way programs had an idea about what they might be rated on and how it felt for them. The fact that they contributed to the construction of the scale may also increase a sense of responsiveness and confidence. Thus they would be more likely to participate.

Another effort to promote buy-in was made by supporting programs throughout the process. After SPARK was launched, consultants were in place to help the programs understand the purpose of the system and orient programs to apply for a rating. They used the scale as a guidance to prepare for the rating. Successfully certified programs were invited to discuss their experience with programs that just entered. Assessors reviewed

the report with programs and guided them to improve. In addition, all the staff were trained to be warm and nice when interacting with programs:

So you must be very inviting, very nurturing, and you come across as very friendly people who can guide you. We provide guidance . . . we coached the assessors to be friendly we want to build the reputation that we are professional and we are friendly and we come across as very nurturing people. That's very important. (admin-3)

SPARK had a good start in setting the environment as hospitable and supportive as possible. They established feedback loops among leadership groups as well as with programs. They also intentionally promoted buy-in in programs. They were also flexible in adjusting policies and strategies to better serve the success of the system.

Alignment with external resources—Systems interventions. Systems interventions, as one of the components of the organization driver, are parallel to facilitative administration. Facilitative administration focuses on building a supportive system environment, while system intervention concentrates on supporting this attempt by aligning and communicating with external agencies and recourses.

Evidence for systems intervention is the committee meeting that is held once a month. The leadership teams are from different departments in education. The purpose of the meeting, as aforementioned, is to raise questions, affirm, and provide suggestions for assessment reports. The fostering of this committee increased communication effectiveness among stakeholders, and also promoted the leadership role that the quality assurance department took in the realm of quality rating and improvement. This meeting may also foster responsiveness in the other leadership groups in supporting SPARK.

The alignment of standards between SPARK and the Singapore curriculum framework demonstrates the connection that is made intentionally with the national standards and the attempt to align each component of early childhood education. However, there is no attempt in working with research institutions other than to provide them with part of the data for research purpose only.

Inform decisions—Decision-making data system. After five years of implementation, now SPARK has “a lot of assessment data.” The administrator was very excited that a big review is coming along soon and she was eager to see how the system has worked up to now. In fact, an annual check also took place each year during the first five years. In the big review, one of the goals was to run reliability and validity of the scale again. “This one (referring to the reliability table on the scale), to me, [is] not very comfortable because the quality was very low at that time, so they are very reliable at emerging level” (admin-3).

Since the validity and reliability scores were generated from the pilot test with only 50 programs (10% of the total number of kindergartens), and the variation was small due to the similar rating each program received at that time; the new examination of validity and reliability this time will certainly be more accurate. As aforementioned, this analysis work will only be done by the agency.

Another goal is to revise the scale to fit better with childcare: “When we developed this (refers to the instrument QRS) we did not consider childcare centers. So certain areas may not fit. In our review study, we want to know what are the items that do not fit, I need to know so that we can refine the scale” (admin-3).

Within the organization driver, SPARK has done a good job in providing a hospitable environment by multiple actions, such as promoting buy-in, fostering feedback loops among leaders as well as between government and programs, and making appropriate adjustments. SPARK also demonstrated its capability in data utilization, for example, using pilot data to inform measurement development, and the prospective five-year review. However, SPARK did not manage to show great progress in system intervention, such as cooperating and making use of external agencies, such as research institutions or institutions that may fund the project.

Leadership.

Revolution—Adaptive leadership. Adaptive leadership is seen through the development and exploration of the new system as well as the new measurement tool. In the three field tests changes were made on the scale itself, the procedure of the assessment, and the whole system. After the launch of SPARK, the replacement of checking assessors' reliability once a year to requiring standardized training, namely Standardization of SPARK Assessment, was another big decision that was made. The cost can be huge if it is not been appropriately approached. In the future, the administrator spoke about making changes on the scale to make it fit better with educational classrooms in childcare centers:

The instrument originally is developed for kindergartens because it is for assessing the education program which the kindergartens provide. Subsequently, the government has decided that SPARK will include the childcare centers.
(admin-3)

Whenever adaptive leadership is observed there must be a corresponding adjustment in policy and practice within the system. These changes should also be counted as part of the revolutionary change. However there are some problems that are easier and more straightforward to address. In those cases, technical leadership is required.

Problem solving—Technical leadership. Compared to adaptive leadership, technical leadership requires fewer resources, and has limited influence on the system as a whole. One example is that during the committee meeting, people found that the report file was getting thicker and thicker, that was when they decided to use an iPad instead of paper for the report (admin-3). Technical leadership may be required on a daily base. It keeps the system's long-term well-being.

Integration and compensation. Despite the fact that SPARK has a relatively short history, it does present a good example of what each driver of implementation science may look like for a system like this. When examined together it is obvious that strong leadership is the key in driving this system. The strategy for staff selection and training is appropriate in Singapore's context. A hospitable and supportive administration environment is necessary for leadership to be effective. The alignment with external resources, however, is not very strong, despite the committee meeting every month. A well utilized data system is in place to provide solid evidence for every decision that was made.

Implementation Effect of SPARK

Over the last five years, SPARK was successfully implemented and appears to be functioning effectively according to the original goals of the system. Since its implementation, more than 400 programs or 1 in 4 programs have been certified. Some of the programs were recertified in 2013. More programs are applying for SPARK. A new higher level of quality appears to be emerging. It is intriguing to see how practitioners view this system and what kind of experience they have had with SPARK. In the following section two parts of the results will be presented. The first part is from teacher survey. This part of results will demonstrate quantitative descriptive figures that describe teachers' overall knowledge, satisfaction, and feelings toward SPARK. The second part of the result is from interviews with directors and teachers. It will illustrate the details of their experiences with SPARK.

Survey results.

Descriptive results. Fourteen participants answered the survey (Table 3). The first question was about their knowledge about a series of components of SPARK. The mean rating for knowledge shows that the teachers had moderate to high amount of knowledge about SPARK (3.86/5). Among all the components of SPARK, most teachers thought they were very familiar with the requirement for teacher-child ratio (4.00/5), followed by teacher qualification (3.93/5) and teacher professional development (3.64/5). The lowest average ratings were for their knowledge about financial incentives (3.00/5).

In the second part of the survey, teachers answered questions regarding their satisfaction toward SPARK. Overall this part of questionnaire received similar scores

compared to the questions about knowledge. The most satisfied component was the requirement for teacher qualification (4.00/5), followed by feedback (3.92/5) and professional development opportunities (3.77/5). The least satisfied component was the financial incentives (3.42/5).

Teachers were also asked to answer questions regarding parents' knowledge about SPARK from their perspectives. This part of survey received low to moderate scores overall. The highest score was parents' knowledge about the rating level of the program (2.93/5), followed by teacher-child ratio requirement (2.71/5). The lowest score of this domain was parents' knowledge about the self-appraisal process (2.22/5).

Teachers were also asked to rate their general feelings toward SPARK. In the first question: to what degree is their daily work being affected by SPARK, the average score was 3.36 out of 5. Teachers' ratings were higher for the second question: how effective is SPARK in enhancing their program quality (3.86/5). Teachers did not express high levels of stress because of SPARK (3.00/5).

Most teachers indicated their program rating was moderately accurate (3.71/5), and viewed professional development as "effective" in improving program quality (4/5). When asking about each type of professional development opportunity, fourteen teachers had attended workshops or meetings, 12 had experienced sharing resources with peers, colleagues or other programs, eight teachers had continuous education, and three had one-to-one coaching experience. Workshop or meetings experience received the highest score in its effectiveness (3.91/5), followed by one-to-one coach mentoring (3.90/5). Sharing resources with others got a 3.75, and continuous education only had 3.56.

Correlations. There were significant correlations between satisfaction toward SPARK and teachers' feelings about SPARK's effectiveness and the pressure level SPARK created (Table 4). Specifically, a high satisfaction in self-appraisal was negatively related to the perceived level of being affected by SPARK ($r = -.59, p = .04$), positively related to the perceived positive effect SPARK has for the program ($r = .63, p = .03$), and negatively related to the perceived stress SPARK created ($r = -.77, p < .01$). Satisfaction with the outsider assessment process was also negatively related to perceived stress ($r = -.64, p = .03$). Satisfaction for professional development opportunities was negatively related to both perceived level of being affected by SPARK ($r = -.61, p = .03$) and perceived stress level ($r = -.67, p = .01$).

Director experience. Participants were recruited from two programs (Table 1). Compared to a childcare center, kindergarten in Singapore is more like a school experience in terms of time schedules. Children spend three hours in the morning and afternoon respectively. It also has a heavier focus on education. However its philosophy still leans more toward early childhood education and developmentally appropriate practice. Program D applied for SPARK in 2011, the very first year it was launched. In addition, it was selected as a program in field tests for SPARK.

The two directors that were interviewed had quite different experiences with the QRIS. This may due to reasons such as program type, time applying for SPARK, and program resources. However, they shared more similar than different experiences and their feelings about SPARK were consistent. They both experienced a sense of anxiety when SPARK was first discussed. They both felt that mentally preparing for it was

important in addition to all the other material logistical preparations. They also experienced a change in their view after being through SPARK themselves, and they were both excited about the future with SPARK.

Fearfulness. The preparation for SPARK started as soon as the system was developed. Word started to spread out that a new system was being developed and everyone knew it would have great impact on them. Some directors did not quite understand the notion of quality assessment and ratings at the beginning because they thought it made no sense to rate their programs like hotels (admin-3). That is also why SPARK is wary of using stars to rate programs. As part of the SPARK assessment, when the program directors being told that there would be two interviews, one with the director, the other with the teachers, directors reacted strongly because they thought it was like “checking on” them with teachers (admin-3).

For directors the feeling of being “frightened” and experiencing “anxiety” dominated the initial stage of SPARK (director C; director D). One director told me that she was always sitting on the bench throughout the workshops and listening very carefully to, not only the trainer, but also to how others talked about the system.

Because a lot of things that we are stepping into . . . the most challenging was the first part, you know. It is the first part. It’s making [an] effort to go in and [experiencing] a lot of unknown[s], you know, and a lot of whether you are actually meeting the standards or not. (Director C)

The panic also came from programs who had been through this process. Program C did not apply for a rating until 2014 because it was a new program. Other than learning

from trainings and workshops about SPARK they acquired quite a lot of information from other programs as well.

Before I join SPARK I hear a lot of nasty things about SPARK . . . The assessors were very fear[ed] . . . They ask questions [and] the teachers feel frozen, [they] don't know what to answer you know. And they are very strict particular[ly] about certain thing[s]. So after [I] hear all the stories [I] will be scared. (director C)

Fearfulness dominated directors' feelings about SPARK at the very beginning. It was mostly the fearfulness of uncertainties this new system may bring.

Challenges. For the most part, both programs had a challenging time with material preparation, including curriculum materials, feedback from parents, and minutes of meetings. "Documentation" became a big word when talking about challenges. Program C did not have a solid curriculum for teaching because it was still new by the time it got evaluated. However, curriculum planning plays a big part in the SPARK evaluation. Thus the most challenging thing for programs is to prepare a curriculum. "The most difficult part for this SPARK journey actually, for me is at first we have this curriculum, we were not, you know, the curriculum was not ready yet" (admin-3). Teachers were not "used to writ[ing] lesson plan[s]" (admin-3), but eventually they received support from the government to finalize a curriculum for everyone.

The other program, program D, also had problems in documentation. Even though they had a solid curriculum in place, director D found that they were losing track of what they had done. An example she gave was that when SPARK came they were looking for very specific documents, like parents feedback. However those documents were not

stored for a long time. They were reviewed and changes were made, and the paper-based feedback was thrown away. Documents like this included meeting minutes, the self-appraisal process, activity notes, etc. Some of the documents were required for “planning,” which is one of the indicators on scale.

One of the other challenging parts for program D was “strategic planning,” which requires centers to: (a) practice annual self-appraisal, (b) have short-term and long-term goals that are aligned to the vision, mission and core values or philosophy and core values, and (c) have center leaders carry out annual action planning (Office of Planning, Research and Evaluation, 2010). Program D talked about even though they have done all of those, it was hard to remember to record everything of it. “[The] issue was documentation . . . early childhood teachers like to do [things], but they find it very difficult to document” (director D). The storage of all the documents also became a problem gradually as the documents accumulated. So program D decided to scan every paper and store it electronically.

Another challenge for program D was to build relationships with elementary schools, which is a quality indicator in the scale.

I can manage them, but to say I want to have a collaboration with the primary school and the primary school do not want to open their doors is very challenging. I can knock and knock how hard, and many places you knock not all want to open their doors and for me it is extra challenging because the school, our school is not in a housing estate, so I don’t feed into one primary school or two primary schools. My children go to, at least 30 primary schools all over the place. (director D)

With challenges the program leaders actually helped the program identify problems and promoted their action in fulfill that part, such as through documentation and curriculum plan. It also reminded programs of some problems that cannot easily, however important, be solved.

Supports. There is a great need for support for programs because SPARK is challenging their way of viewing this profession and the way they organize and process information. In the interview directors talked about different kinds of support they received, such as training, coaching, parents and community support, and support for teachers. One administrator told me that they used every way to help the programs understand the system. This idea was confirmed with the feedback from programs.

We went through a lot of training. In our . . . organization we have [them] conduct training for us, we also have the help of our district head who . . . actually train[ed] in this SPARK instrument, so they also came and guided us. Yeah. And besides that, MOE (Ministry of Education) . . . also conducted all these self-appraisal SPARK training[s] and I also attended that training session. (director C)

It has also been helpful for large programs like program D that has many teachers and find it hard to coordinate everyone's time, to be able to invite SPARK trainers to their center to do training with all the staff (Director D). Being accessible is even more frequently mentioned than being flexible. One of the supports that programs receive is from the district head or an immediate supervisor. These people have direct contact with MOE and have more information and knowledge about SPARK than program directors do. Strong feelings were expressed by director C: "Every time we do this and then I have meetings with my teachers and certain things I'm not too sure [about] so she will say

“you need me to come down?,” “okay”; we just have [to] call . . . and we have session. And that was very, very inspiring” (director C).

Director D was especially grateful for her teachers who were very supportive throughout the process. She talked about how anxious she was because her center was one of the programs that got evaluated in the first round. As she talked to the teachers she found them to be open-minded and supportive for her to apply for SPARK: “And our teachers, we are blessed. Our teachers are receptive. ‘It is okay, let us change, it is okay, let us change.’ So that receptiveness is very difficult to find” (director D).

The program administrators also benefited from the sharing of experiences by those who have gone through the process: “So we are going to visit and we have a chance to interact and see what other things they need to prepare, you know. How they move on. This is very, very useful. We went [to a] few of the centers like this SPARK [who] just got the certification” (director C).

The government has done a lot to help programs become familiar with and get used to this new system. They also support programs to improve through reviewing the assessment report with them. However, other than that, SPARK is not able to provide further support for a program’s improvement. Take program D as an example. They have difficulty making connections with elementary schools that their graduates will attend. What the system could do is to provide a list of potential elementary schools for their reference. However, no more support is offered beyond that (director D). Moreover, other than assessing professional development plans of a center, and providing a small amount

of scholarship for professional development (admin-3) there was not much other support available for programs or teachers to make improvements.

Support in the form of instruction and guidance are abundant from the government's end. Directors were also grateful for the inside support from staff in the programs. However less support could be seen in the actual quality improvement part, either financial or policy-wise.

Mind switching. A switch in directors' perspective was evident after they experienced the QRIS process. This was especially salient for director C because she applied later and knew about the mixed reactions to SPARK from other programs. There was first a switch in her fearfulness in the assessors, who she heard were strict and inflexible from someone she talked to.

But having gone through to this journey it's totally different from what I hear. Assessors are so friendly. [If] they ask you a lot of things . . . and you don't understand, they rephrase it. So actually I find that they are ok. They are not as fearful as what we heard of. And they came to the center, they came to assess my teachers, my teachers felt very comfortable. They don't show that they are arrogant. They just came to see whether you know. (director C)

And then she found an unexpected smoothness of the preparation for SPARK. "I heard people say that they have to stay up into the middle of the night and do all the paper work. I don't see that in me" (director C).

She attributed this smoothness to mentally preparing the teachers as well as she had in the preparation phase. Director D also talked about similar feelings that it was a relief it went so smoothly even though there are still problems to be settled. These feelings were just what SPARK aimed for at the beginning stage (admin-3). Great efforts

were made on the government's side to make the process smooth and friendly for programs.

Motivations and hope. The words used by directors to describe their feelings for the next assessment three years later were very different from those used in describing their feelings at the beginning stage. Directors found it worthy to go on this journey. They were looking forward to the next assessment, even though there is still going to be anxiety involved because they will be evaluated on higher standards. Program C also talked about the process of waiting for the certification as “fun and excited” when she looked back. She found it was like “baby delivering,” for which you did not know the gender of the baby but you were so excited to see the baby. When they knew they got the certification the “teachers screamed.”

Teacher experience.

Mixed feelings. Compared to directors, teachers had mixed feelings toward SPARK. Teachers may experience less mental stress than directors do. This may due to the nature of the assessment, which puts more weight on the administration side. However, teachers are the agent of a program; they are directly involved in program activities. In addition to the feelings of fearfulness about a new process, some teachers did not feel comfortable with the idea of being assessed with a certain standard. Teacher G had worked in early childhood education programs for years. She was also a team leader in that center. She knew that some teachers, including herself, did not buy this assessment idea.

I mean [we] are we really giving them quality education, I mean for me I feel personally quality should come from the heart, not from skills, or you know, this kind of system. But then on seeing that to run a center, and I think to be recognize[d] I think this kind of system is necessary, to [a] certain extent. But we should not be falling into that, you know? (teacher G)

This teacher and the teachers she represents were suspicious about the system's philosophy. They felt that quality was not measurable, it should be teachers' affection and passion that defines the quality, not something that is visible and concrete. Even after she changed her mind of thinking that this system is entirely meaningless to it is "necessary," she still kept a certain degree of suspicion and cautions toward not "falling into SPARK."

Contradictory to teacher G, teacher E was a new teacher in the ECE field, and she recalled her feeling about SPARK when she knew they were going to apply for it as "curious." "I [was] actually kind of a bit curious [laughing] . . . just a little bit curious. And I wanted to go through it. Yeah. I want[ed] to experience it. I guess it will turn out fine, it is curious" (teacher E).

It was easy to tell that she was excited about the new system because she was curious about it after hearing so many things about it from others.

Confidence and morality. Despite the mixed feelings, most of the teachers were confident in getting through the assessment. Teacher F's confidence came from an understanding of and a trust toward SPARK:

But I think if we are all in the same boat and we . . . understand the importance of this QRIS and the benefits [there] will be for the children and the centers as well, I was okay with that too. And then knowing that [we are not] doing [this] by

ourselves, we were a group of educators that would like to move forward, right?
(teacher F)

This confidence has also been through a developmental process. Over the two years of the QRIS development, the program that teacher F was in was invited to do pilot testing. They also received training and coaching as other programs did. The idea of a quality rating system gradually was accepted by most of the teachers in that program. They mentioned “the journey” several times, which may mean that this value is now well accepted and internalized as a belief in the system. “We were involved in the focus group discussion so it was about two-year process to actually understand [the] quality rating scale, [then] accept the fact that it is [a] very important factor for the entire early childhood education in Singapore” (teacher F).

Teacher D and teacher E’s confidence were derived from a trust toward their program and what they do on daily basis. They felt what they were doing everyday was correct, so the result couldn’t be too bad: “Because we put in effort in it . . . we also know that ‘ok, we are not so bad’ . . . I think we should be okay” (teacher D).

Challenge. The biggest challenge, according to both teacher F and G, was to prepare documents appropriately. The first challenge one teacher had was to prepare the documents. As was also been mentioned by administrator D, they did not have a habit of keeping and storing materials. Although they have done a lot activities, no evidence could be found when came to the assessment. For teachers there was also an uncertainty about the appropriateness of the lesson plans. Since the program they were in was a big

program, teacher G, as the team leader of a learning group, had to make sure that the lesson plans were well-written and properly kept.

Another challenge expressed by the teachers was to prepare people. Teacher F was also in a management role in that program. She found it was challenging to promote “buy-in” with teachers at the beginning. She wondered “how to work with the teachers and how to convince them [of] the benefits and how to go about it and highlight it to them that this is how we will go forward” (teacher F). When they decided to apply for SPARK several learning groups were formed. A group or team leader was selected to lead that learning group to make changes. Teacher F found it was an important yet challenging task for her to plan and organize the groups, to select that leader who was expected to be able to work with team members “with the right knowledge and the right attitudes and aptitudes” (teacher F).

We don’t want somebody’s like racing against each other. We have to look at that carefully. Some teachers raised a voice, “Why would you put me in this particular group?” We have to have the right reasons to make sure that teachers can accept that reasoning behind it. (teacher F)

Teacher G was challenged when she tried to change the way of working. Not only did the work load increase, more importantly, she was responsible for helping teachers in her team change their way of working to meet the criteria of SPARK: “I have to . . . guide them, to have . . . [a] formalize[d] a lesson plan, computerize it because they used to do and return. So there were many changes for them, so we have to . . . [go] slowly, you know, do it, and then they were willing to . . . do it” (teacher G).

Other challenges came from the understanding of the scale (teacher D and E). There were some terms need further definition or more examples to help illustrate key ideas. However, when this happened, teachers could always reach their director and the director could find someone to help.

Program-family connections. Thanks to the advocacy from the government as well as from programs, when SPARK was launched many “educated parents” were aware of it. According to the survey results, parents’ knowledge about SPARK received an average of 2–3 from teachers’ perspectives. Teachers as well as directors thought that parents had limited knowledge about SPARK, other than knowing that they are going through an assessment process and understanding what the banner of SPARK outside a program meant.

Despite limited knowledge, parents did cooperate while the assessment was going on. They tried to be on schedule when dropping their child to make the transitions easier for teachers (director C). Some of the parents even came to help while teachers were busy preparing for SPARK (teacher F). They sometimes worked one-on-one with a child in a learning group, and they come regularly. They even celebrated with teachers and directors for passing (director C). However, when it comes to choosing programs, parents do not give programs with SPARK certification more credit. There are other more realistic things to consider, such as the distance to commute and the price “I never heard people, ‘I came here because you got SPARK’” (teacher F).

Family/community connection was what program D most worried about. As aforementioned, this program was not in a housing area which means they did not have a

natural connection with communities or families around it. Most families chose it is due to its reputation and a national language environment it can provide (director D).

Value of SPARK. The value of SPARK was well appreciated by all the participants. Director D viewed it as guidance for their program and helped them to narrow their focus in order to have the greatest benefit for children:

The understanding of the different trends (is most important). You see, sometimes we think we are here, we are [at a] performing level, but actually we are [at an] emerging level. Because our understanding of the trend is not clear. So the clearer the teachers and the leadership is, about the different trends and the level, we can have an understanding of what the next level is and therefore we can work towards that level. For me that is important. (director D)

She also thought it was an affirmation for what they have done. “For me SPARK is affirming my practices, affirming my understanding of quality” (director D).

For director C, SPARK is a solid proof for the quality of her program:

It’s not an exam, yeah. But it is like a [way to] tell people that all your center has all the processes in place. Whatever we do that is the accountability and it’s [what] we are doing it [is] transparent. And we met a lot of quality [standards] that [are] needed. It is all about how we deliver our lesson and I think SPARK is all about . . . bringing the standard up, ECE standard up. (director C)

Teachers seem to prefer to view it as a “learning process for everyone.” “It’s an eye-opener that’s one of the area[s] that we have to document children’s learning at this age because there’s not test or assessment or grading, not like in primary school. That’s part of the teachers’ professional development” (teacher F).

Teacher F felt that it did not promote competition but it let people know where each program’s quality was. Teacher D and E approached the value question from

children's and the nation's perspective. They thought it would bring an ultimate benefit to children and as a result to the nation.

The effect of SPARK's implementation on programs was mostly positive. Both directors and teachers demonstrated an understanding of the purpose and philosophy of SPARK and expressed certain degrees of appreciation for this process. This can be viewed as one of the dimensions of implementation fidelity in that SPARK has the potential to move on to a new phase because it demonstrated high quality itself and will very likely to be supported by programs.

Implementation Stages of SPARK

Since SPARK has a relatively short history, its developmental track is clear and linear when looking back. It took a few years for the leaders who were in charge of developing a quality rating and improvement system to explore ideas and develop processes before SPARK was launched. In 2008 the administrator started to pay visits to different countries where QRISs were implemented for years. The most salient characteristic of exploration stage is "frequent, iterative interactions among and between stakeholder groups, knowledgeable experts, and perhaps other sites or organizations that have already implemented the new way of work" (Metz et al., 2013, p. 26). Those visits led to the construction of the measurement tool, the Quality Rating Scale (QRS). During the exploration stage, three field tests were run to further test the measure.

There were times when the development group needed to go back and update the measure. For example, in the process of writing the measure they found that a guiding principle was needed, so they went back and added that part. However, the measurement

tool is not all there is to the system. Other aspects like staff selection and training, the formation of leadership and committees, and the adjustment needed in policies were put into action. There is not a visible cut point between the exploration and installation stage for SPARK. It seems that the day that SPARK was launched could be viewed as the cut point. However even before SPARK was launched officially, actions were taken to promote buy-in among stakeholders. Examples are inviting programs to participate in the measurement construction, and advocating for the new system in public long before it was launched. The new practice was initiated with programs that applied to be assessed.

The development of practitioners' confidence and competence is one of the most important features of the installation stage, and it has not stopped since SPARK was launched. Assessors are trained to be warm and encouraging in the interactions with programs. However, according to director C who learned from some programs that had experienced the evaluation that the assessors were initially not as friendly and the process was frightening. It is assumed that in the early installation phase, the team did not realize how important it was for assessors to be encouraging and friendly to programs. Some assessors were not aware that coming across as authoritarian may reduce programs' motivation and willingness to participate. However, the experience the directors and teachers from both programs had was positive: The assessors were nice and flexible in their schedule throughout assessment and report review phases. It could be that the training was intentionally adjusted to fix the initial "frightening" impression. The promotion of buy-in continues to be the key feature even in initial implementation stage.

To promote buy-in does not only require assessors or consultants to be welcoming and encouraging it also requires them to be experts in the early childhood education field. That is why assessors who did not have an ECE degree were sent to pursue a diploma in ECE. This was done in the exploration stage. Sustainability is not a certain stage, rather it follows every stage. To sustain the system's fidelity it is important to check and reassure assessors' reliability and reexamine the alignment between policy and practice regularly. It seems that SPARK has successfully installed the major components and has the key features in implementation science. The primary job of it is to sustain fidelity, continually promote buy-in, and move to the next stage. Thus, the current stage that SPARK is in should be the initial implementation. After five years of implementation SPARK is at the point where administrators may need to examine the effect of implementation and to re-test validity and reliability of the scale. Over the five years problems have emerged and most were solved along the way. For instance, the reliability check process has been changed to a standardized training. Furthermore, EC experts were invited from the U.S. to train both new and old assessors. The most urgent mission for SPARK is to develop criteria for the highest quality level, Accreditation, since programs keep reaching toward the other two quality levels. This continuous improvement corresponds to what typically happens in the initial implementation stage.

When examining each driver of the framework, the competency driver is almost fully implemented since there are effective staff selection and training protocols for assessors and consultants. According to the administrator, the selection and training are rigorous. In addition to 100 hours' instrument training, prospective assessors will spend

several days in the field to practice and do reliability checks. For consultants who are already ECE experts they also receive instrument training to be able to assist programs. The reason that it is still not fully implemented is the lack of a fidelity test based on a large pool of data that are collected through the years. Due to the shortage in manpower, the regular reliability check has been replaced by standardized training twice a year. Little is known about the fidelity of this new strategy. Upon proving of the fidelity of this strategy we can say that competency driver is fully implemented.

In terms of the organization driver, it is impressive to see how effective the leadership team was in the past few years to develop SPARK from nothing to its current stage. A hospitable environment was created through the years by the leadership. The quality assurance department of this team is in charge of the integration and monitoring of the whole system. Communication within the team was quite effective as well since several departments cooperated on measurement development and report review process. Adjustments were made in a timely manner. This was based on the identification of barriers and capability in solving them. The shortage within the organization driver involves its lack of connection to external resources and institutions. This does not necessarily bring disadvantages to the system given Singapore's unique cultural and political environment; however, special attention needs to be paid to this feature of the system.

Beijing

Components of Beijing's QRIS

In 1989 the Beijing Kindergarten Quality Rating System was launched. Thus, the quality rating system has been implemented for decades in Beijing. Even though the same standards apply in all districts (16 districts in Beijing) there is a wide range of variability in each district in terms of implementation. In this study only one district of Beijing, Xicheng district, was under investigation. Xicheng district was picked by recommendation. Professionals considered it as the most advanced district in implementing QRIS in Beijing.

The most salient components of QRIS in the Xicheng district are the standards, measurement, and improvement process. There is a clear absence of financial incentives and consumer education. The evaluation of quality is partly based on the standards, but there is not a measurement tool to quantify the results. There are multiple ways assessment and improvement are largely interwoven with each other. In this section results related to QRIS components from the interviews as well as the document review will be presented.

Standards. Quality rating standards are based on the *Regulation* and *Guideline* documents. There are 10 criteria for structural quality and 26 criteria for process quality. Structural quality includes the physical environment, such as size of building, equipment and materials, and staff qualification. Process quality includes four dimensions which are Administration, Education and Pedagogy, Health and Hygiene, and Child Assessment. Structural quality is referred to as “Ranking” (R). There are four rankings (from 1 to 4).

Process quality is referred as “Category” (C). There are three categories (from 1 to 3).

The combination of R and C represents a program’s quality.

In the standards for process quality, Administration is weighted the heaviest, with 10 out of the 26 items. Items in this dimension include program value, program climate, teacher professional development, program management rules, regular assessment, and document management. Education and caring has six items that relate to educational planning, activity implementation, educational environment creation, and parent connection. Health and hygiene is rated on five standards that are related to health examination, recording health and nutrition plans, and disease prevention. Finally, child assessment also includes five items including the assessment of health, intelligence, moral behavior, aesthetic taste, and personalities.

In addition to the standards for most of the programs, there is an add-on rating standard for the highest quality level, which is called the Demonstration level. Besides meeting all the requirements for 1R1C, programs that apply for Demonstration level will be rated on three dimensions, 10 items. This additional standard has a higher requirement on both structural and process quality. For example, it has more specific requirements for staff qualification, including years of experience for directors and teachers, and a higher qualification requirement for a school doctor and health care staff. Demonstration program standards have a specific emphasis on creativity, progressiveness, being research-based, and holding responsibility for outreach in the community. In addition to the high standards that are set for 1R1C, demonstration programs are expected to lead and help other programs. For instance, there are specific criteria such as helping lower-

level programs every year, or establishing a bond with at least one rural childcare and providing a two-month on-site learning opportunity for at least one teacher from that program every year. They are also expected to be responsible for the community as well as for whole society through different kinds of public activities and advocacies.

The standards for all the programs were updated in 2000 and the latest demonstration standards came out in 2005. Both sets of standards weigh administration more than any other domain. According to the administrator (admin-4), the actual assessment was much more intensive than the standards on the paper. Xicheng MOE was in the process of coming up with their own standards for quality rating. When reading through the current standards it is not hard to find that some standards are too vague or too general to measure. The assessment of the classroom focuses more on the environment, equipment, curriculum, and planning, and less on teacher-child interaction quality throughout the day. Child assessment is considered as an important indicator of program quality.

Measurement. The quality rating in Xicheng is unique. As mentioned, there are two rating systems, one is for Demonstration programs; the other for the remaining programs. Groups that are composed of several programs and are led by a Demonstration program are formed based on their geographical location. These groups are called “Learning Communities.” Within a group there are several programs that are at different rating levels (admin 4). Multiple assessors will do the assessment on the evaluation day. The assessors are also experts in a certain domain. Experts in health and hygiene will evaluate health and hygiene in the program, experts in curriculum will look at lesson plan

and lesson implementation, administrative leaders will be in charge of document review and interviews with directors about the administration and management in the program.

In the first half of a year, programs that apply for Demonstration or programs that are already at the Demonstration level are evaluated by a group of experts from all over the city, excluding experts from Xicheng district. On the evaluation day, representatives from other programs of the same group will also be present to learn from the process. In the second half of the year, programs other than the Demonstration programs will be evaluated by leaders from Demonstration programs, who are also considered as experts in a certain domain. The experts that assess the Demonstration programs will also be present, but they will not do the actual scoring. They will only be instructors or coaches on that day.

The measurement is not a holistic scale like ECERS-R or QRS, it is composed of several scoring sheets that corresponds to the standards. The measurements are aligned with each standard and reflect each item in the standard. There are very detailed criteria for each item. For example, for structure quality, such as the size of space, the number of toys or books for per child in the classroom, and staff qualification are quantified and specified. The scoring sheet for Education and Pedagogy includes 107 items over three dimensions: Curriculum implementation, teacher and child assessment, and lesson plans. This part of the assessment is based on observation. Interview and document review take place after the class observation. However, the content of the evaluation, according to admin-4, is broader and more intensive than the actual rating sheet. That extended part is not written in words yet. They seem to be based more on the experts' experiences.

Improvement. Improvement is embedded in the evaluation. On the assessment day, experts or assessors make comments and give suggestions to program leaders and teachers as the assessment goes forward. The assessors also have roles as coaches and technical assistant providers at the same time when they come to a program. They are the authority figures to the programs because they are considered “experts” in a certain field. However the assessment day is viewed as somewhat scattered and limited since it is based on experts’ immediate reaction to the situation. Furthermore, according to the administrator, the suggestions in the final report are not always supported by rational or concrete evidence (admin-4). They may be very general and too vague for programs to actually make changes. The formation of the learning community, however, does provides more opportunities for programs that are not yet at the Demonstration level to learn from the Demonstration programs and to acquire more hands-on experience when they are evaluated. The administrator proudly called it a “face-to-face evaluation” instead of a “back-to-back” evaluation as it was in the past. Since the same Demonstration program comes to evaluate each year and they are familiar with the condition and status of each program, they can easily tell what is still missing and what has changed in that program. It also provides the Demonstration program with a sense of achievement because they see the positive changes that are based on their suggestions. The sense of achievement turned out to be a strong motivator for them to be the assessors and professionals in this system.

Other support comes from professional development. All the professional development opportunities are provided by the government. There is no fee for teachers

or programs to participate in any training. However, opportunities are assigned by the government as well. Demonstration programs get more opportunities than other lower-level programs. It is considered a reward for a high quality program. It also serve a certain function that the participants should be qualified to conduct training in the learning community based on what they learn. Thus teachers in Demonstration programs are considered privileged when many “good” opportunities come along (admin-4; teacher N). Rating results are directly related to professional development opportunities. The alignment between rating and professional development is so strong that it could be viewed as a rewarding policy.

Financial support. Financial support is not associated with quality rating in Xicheng district. There used to be a small amount of financial compensation for Demonstration program leaders for their evaluation job. However, due to a national policy that requires the ‘transparency of a public servant’s income,’ this financial compensation was no longer allowed. Thus, for this legal reason only a certificate of merit is awarded to Demonstration program leaders for their hard work in evaluating and helping their sister programs.

Consumer Education

Consumer education is not considered one of the components of KQRS. The “government is not obligated to advocate this notion of what KQRS is” (admin-4). However, after decades of implementation, both the administrator and program directors thought that parents were aware of the program’s rating (admin-4; director F). “They

(parents) know that 1R1C means the best, but they may not know how many different combination[s] of ratings there are” (director F).

In KQRS, standards are embedded in measurement, and improvement is embedded in assessment. In fact there is no clear cut line between any two of the components. The content of assessment is beyond measurement, however it aligns with standards. The improvement is aligned with the assessment results however it may not merely be based on assessment results. The notion of community and social responsibility is one of the fundamental values that guide the development of KQRS in Xicheng district. To be specific, the formation of the “Learning Community” is the best representation of community and collective society. Programs in the community engage in joint activities, including quality rating and improvement. It is also considered as a compulsory duty for the best programs, demonstration programs, to guide the rest of programs in the same community.

QRIS’s Implementation in Xicheng, Beijing

Quality rating in Xicheng is called a “three-level rating,” which refers to three different rating types. The first type is self-appraisal, which is required for every program each year. The second type is group rating, which is conducted by Demonstration program leaders. The third type of rating is expert rating, which is conducted by experts from different areas. Group rating is only for programs that are lower than the Demonstration level, while the expert rating is only for programs that apply for the Demonstration level or are already at the Demonstration level. However, on the evaluation day, all the parties will be present. For example, when assessing a

Demonstration program, leaders from all the sister programs in the same learning community should be present in addition to the assessors (i.e., experts). The same applies to the assessment of non-Demonstration programs: While the leaders of Demonstration programs are evaluating that program, experts should be present to give suggestions and instructions.

The annual assessment as well as the group assessment has been implemented since 2006 when an act was issued by Xicheng MOE. This act is called a “Learning Community.” In this act programs are assigned groups (i.e., communities) based on their location and quality level. The tenet of this act is to promote mutual learning within learning communities. “Every assessment is a specific case for every program in that community. They learn from each assessment not merely from their own assessment” (admin-4). According to the administrator, the implementation of this new system significantly brought up the overall quality level in this district from 40% to 70% of 1R1C programs. It also enhanced the connection among programs (admin-4). While keeping many traditional strategies, such as hiring experts as assessors and embedding improvement into assessment, efforts were made to incorporate many new strategies into the existing system.

Competency driver.

Staff selection. The “assessors” in the Xicheng KQRS are not formally trained as assessors. They are either experts in a certain field (e.g., curriculum, health and hygiene, administration) or the directors or group leaders from Demonstration programs. Some of

the experts are retired principals or department leaders who have specialties in a certain or general domains:

The experts and well-known principals will conduct the rating based on the quality level ([they] use different standards). For example, experts from management will evaluate program management, [a] research expert [will] evaluate [the] educational research domain. These experts are selected from every district and county by Beijing MOE. Experts will not evaluate program[s] in the district that they are from. (director G)

According to the administrator, some experts were in many different education-related occupations before they retired and became assessors. They are hired through the district MOE and are recruited from all over the city. Experts from a certain district are not eligible to evaluate programs in that specific district. Other than assessors, the trainers for teachers' professional development are also considered as a part of the system. The institution that provides training for teachers' professional development is called the Education Institution. The trainers are selected from schools or childcare centers. They were the most experienced and knowledgeable teachers when they were in the schools. This is considered as a promotion for teachers. Besides training, the trainers also engage in educational research (admin-4).

Staff training. A systematic training for assessors is absent, since all of the assessors are considered to be experts in early childhood education field. The training for trainers is done within the institution through assigning them research tasks. This is considered a way to maintain their professionalism (admin-4). Staff in Xicheng's KQRS system are mostly teachers or school directors who grow and develop to be assessors or trainers.

Organization driver.

Setting the environment—Facilitative administration. Facilitative administration refers to a series of practices that set the environment for a system's installation and development. A coalition between Xicheng district and Xuanwu district took place in 2010 and it became a single district which took the name Xicheng. There used to be two administrators in the MOE for early childhood education. The north Xicheng (i.e., previous Xuanwu) and the south Xicheng (i.e., the original Xicheng) were not merged in terms of their policies regarding KQRS until 2015. In 2015 the present administrator, who used to be the administrator for the original Xicheng district, took charge of both north and south Xicheng's early childhood education. Since then she started to import the new system that has been implemented in the south Xicheng since 2006 to the north part of Xicheng. "It was pretty painful" said the administrator. "It is not to say that their system is not good. It is just that they did not include all types of programs in the quality rating system." It is important to make sure that the leadership is committed to the system. Since the concept of "Learning Community" was brought up by the current administrator there is little doubt that she will try her best to promote the success of this system. In the interview it was clear that she was proud of introducing this new system and optimistic about the new way the system would work in Xicheng. In order to encourage programs to adopt the idea of Learning Community she regularly engaged in the assessment process herself.

The setup of the learning community not only enables communication on the assessment days, but the most important function of the community is to promote regular

communication among programs within a community. “When a program comes up with a research topic other programs will send representatives to study and discuss that topic” (admin-4). Program leaders of Demonstration programs also meet to discuss issues, concerns, and challenges once a year. The communication and feedback loop among programs are built in the system, and it is based on a dynamic learning approach. The communication between programs and MOE is based on both official and unofficial procedures, such as unplanned visits, or phone calls. The relationship between programs and MOE is more like friends and colleagues than strangers or outsiders in the daily communication. However MOE is also treated as an authority by programs.

The creation of the learning community served to decentralize the power and responsibility to programs that were capable of guiding others. It aimed to increase a program’s motivation to enhance the quality. It also took some burden off the government’s shoulder. In this way all types of programs could be included in the rating system. In the old system quality assessment was only done by experts, and “it is sometimes hard for programs that are of low quality to understand some suggestions experts made. The experts made the suggestions and they left. No one is available to give programs further suggestion in how to improve” (admin-4). In the new system, however, since the Demonstration program will evaluate all the other sister programs in the same community, they are familiar with all their situations. They are also located close to each other, which makes it is easier for the sister programs to seek help from Demonstration programs even after the assessment.

In order to create a hospitable environment for programs to participate in quality rating, it is also important to reduce administrative barriers. In Xicheng's case, the resistance was from programs. "It was not an easy task" said the administrator, "some felt that to assess once a year is too much for them. Some Demonstration programs felt it was a great burden to them given that [since] they already had billions of things to deal with." To make the transition smooth, MOE did not force any program to participate at first, but experimented with a few programs and let other programs see the effect. They also hired very strong experts after their retirement to be assessors and instructors. After a time of implementation, positive feedback was received from programs that participated, so more and more programs took part in the new system.

In building a hospitable environment for implementation the leadership was in place and committed to bringing changes to the system. Communication loops were established between programs. Administration barriers, such as program resistance and uncertainty, were mostly cleared by using strategies that promoted buy-in. The inner environment was set. According to the theory it is also important to have support from the external resources.

Alignment with external resources—System intervention. System intervention, as one of the components of organization driver, is parallel to facilitative administration. Recall that facilitative administration focuses on building a supportive systems environment, while system intervention concentrates on supporting this attempt by aligning and communicating with external agencies and resources. Xicheng's MOE promotes leadership by decentralizing the rating responsibility to Demonstration

programs. As aforementioned, some Demonstration programs were against this new policy because it would take away time and energy with no financial reward. However, according to admin-4, when they came back to sister programs and found that they did make improvements based on their evaluation and suggestions, this was rewarding. The Demonstration program leaders acquired a sense of achievement and more motivation to guide the sister programs.

Besides QRIS, there is a separate system called a Supervisor system that monitors program quality. Beijing MOE, rather than Xicheng MOE, operates the Supervisor system, which aims to supervise educational programs' educational quality. The Supervisor evaluation happens every 3–5 years. It evaluates the overarching philosophy of a program and provides directional instruction and advice for program development. Compared to the Supervisor evaluation, QRIS has more assessments of many specific details of a program, but “the tenet of both evaluation are similar” according to admin-4. However, it seems that there is not cooperation between the two systems. In the year that a Supervisor evaluation is expected programs need to prepare for both evaluations.

Rather than nurturing champions, Xicheng's MOE focused more on nurturing a harmonious community. Harmonious society has been advocated by president Hu and the leader team since 2004. It refers to a harmonious, hospitable, and balanced society. The tenet of Learning community corresponds to the idea of harmonious society. Actually even before this concept was well-known and accepted in China, Xicheng MOE started to include all types of programs in the rating system and let high quality programs help them to achieve a higher quality level (admin-4).

It is not right that good education and good quality are only available for some children but not for others. The creation of “Learning Community” is to end this trend and to create equal opportunities for every program, including public, private, and community programs, to develop. (admin-4)

There is no formal data system in Xicheng’s KQRS that informs decision making. Most of the decisions are not based on systematic investigations and data collection but rather on personal experiences and values.

Leadership. The very basic idea of the “Learning Community” derived from the administrator’s value about education. She believed that no one should be left behind. Thus, she was the one who proposed this act and eventually enacted it. It was not an easy task to convince all the stakeholders to accept this new act, according to the administrator. First of all, it is ambiguous whether it is legitimate to have Demonstration programs rate lower level programs given that they are actually at the same level in terms of their power and rights. To enact this act she had to ignore this fact. Second, it is set by law that no financial compensation or reward could be provided to programs or people who did the evaluation. MOE needed to make sure that Demonstration programs accepted this task and were willing to do it, and the sisterhood programs were willing to accept their evaluation.

The administrator is very aware that it is not typical for parallel programs to evaluate each other, and she experienced strong resistance when the idea was first brought up. However, she was determined to take the risk because she foresaw the good outcomes it could bring in the long run. In order to convince programs to participate she

let part of the programs participate first and advocated the fine results among the rest. It worked, at least according to the increased number of high quality programs in Xicheng.

This is an excellent example of how adaptive leadership is important in determining the implementation of a system. It is apparent that the administrator made the most of the decisions by herself. It was efficient and straightforward. However it could be problematic if those decisions were not made based on accumulated and concrete evidence or comprehensive discussions and field testing.

The implementation of KQRS in Xicheng district is featured with its three-level evaluations and the creation of learning communities. The leaders intended to create a harmonious community in which everyone had opportunities to be supported. It is intriguing to see how programs reacted to this new way of work as well as in general to the rating system.

Implementation Effect of KQRS in Xicheng, Beijing

Kindergarten Quality Rating System (KQRS) was installed in 1989 in Beijing. After decades of implementation there are 281 programs rated 1R1C, and 89 rated Demonstration level. In Xicheng, specifically, there are 20 1R1C programs (website, 2014) and 17 Demonstration programs (admin-4). The quality rating was conducted every three years until 2006 when the act was issued. Since then programs have been assessed once a year, regardless of type of the program. This was also when the Learning Community became a part of the system, and the three level rating came into being. Currently the proportion of 1R1C level is 70% in Xicheng, compared to 40% in 2002 (admin-4).

Survey results. Fifty-four participants completed the survey. As recorded in Table 2, most of them were lead teachers ($n = 42$). The first question on the survey asked about their knowledge of a series of components of KQRS. The results suggested that the teachers believed they had a moderate amount of knowledge about KQRS (3.25/5). Among all the components of KQRS, most teachers thought that they were very familiar with the professional development associated with improvement (3.89/5), followed by feedback of assessment (3.65/5) and program self-appraisal (3.44/5). The teachers were the least confident in their knowledge about outsider evaluation (3.09/5).

In the second part of the questionnaire, teachers answered questions regarding their satisfaction toward KQRS. Overall, this part of survey received similar scores compared to the questions about knowledge. The most satisfied components for teachers were professional development opportunities (3.96/5), followed by feedback (3.87/5) and outsider evaluation (3.86/5). The least satisfied component was the financial incentives (3.57/5).

Teachers were also asked to answer questions about parents' knowledge about KQRS from their perspectives. This part of the questionnaire received moderate scores overall. The highest score of this domain was parents' knowledge about the rating level of the program (3.92/5), followed by staff qualification (3.15/5). The lowest score of this domain was parents' knowledge about financial incentives (2.65/5).

Teachers were asked to rate their general feelings toward KQRS. In the first question: To what degree is their daily work being affected by KQRS, the average score was 3.50 out of 5. Teachers rated the second question higher: To what degree do they think

KQRS helped the program improve (4.06/5). Teachers expressed moderate levels of stress due to the KQRS (3.36/5).

Most teachers believed their program rating was accurate (4.54/5). The overall effectiveness of professional development in improving program quality was rated 3.77 out of 5, which falls between “somewhat effective” and “effective.” When asking about each type of professional development opportunity, 49 teachers had continuous education experiences, which means taking course to receive a higher degree, 48 had had workshop or meeting experience, 36 had experience in sharing resources with peers, colleagues or other programs, and only 12 teachers had one-to-one coaching experience. Resource sharing experience received the highest score on the effectiveness in promoting program quality (3.74/5), followed by one-to-one coach experience (3.72/5), despite its low frequency. Continuous education received the same score as one-to-one coaching experience (3.72/5).

Director experience. The three directors who were interviewed came from three different Demonstration kindergartens in Xicheng, Beijing. The three programs are among the most well-known programs, even in Beijing. All three programs have a long history that range from 48 to 112 years, and 8 to 20 classrooms (Table 1).

Motivation. Motivation for getting a higher rating does not come from the hope of receiving more funds or having increased enrollment since, as aforementioned, there are no financial incentives associated with a rating level. Given the large need for child care in cities, programs do not have problems with not having enough children to enroll. Thus, the motivation lies in the reputation, like two of the directors talked about:

When we apply for Demonstration level we were very motivated, since we're already in C1R1 we have to get the Demonstration (or else it will be humiliating [loss of face]). Teachers were very motivated after our director announced we will apply for Demonstration. (director E)

Normalized. The most dominant feeling expressed about assessment was that the standards of assessment had been normalized in the program. All the staff in the program felt they kept on doing what met the high quality standards on a daily basis, so when the assessment team came they did not worry too much about it. In addition, director F felt that the standards on the *Guidance* and *Outline* were clear enough to guide their work. As long as they comprehensively understand the two documents and do whatever aligns with the standards they felt they should be fine in meeting the quality rating standards.

High pressure. Even though directors were mentally prepared for the assessment, they were not free from the heavy workload that was associated with assessment.

Preparing the document became the most energy-taking task for programs,

because it is very difficult to manage all the documents, including meeting minutes, recording, summary, plans for teaching and educational research, class arrangement, conferences, family and program connection, big events, safety, hygiene and health, etc. (director G)

Director G also felt that the leaders experienced more pressure than teachers did because teachers only needed to worry about whatever they were told to do, while leaders needed to consider the whole picture as well as the details. It seems that normalization and high pressure are contrary to each other. However, as director G later pointed out, most of the pressure comes from the comparison with other same level programs, rather than from the assessment itself.

Responsibility. The pressure that directors felt partly came from the responsibility they were taking. When being asked about their general feelings toward assessment, two directors used the word “responsibility.” Director F thought of being assessed as a societal responsibility that programs should take because it is their responsibility to show the parents as well as the public what the quality is. “I think it is the government’s responsibility to monitor us. They are supposed to manage kindergartens. I also feel that it is a good thing for programs to be monitored” (director F).

Demonstration programs also take responsibility for evaluating sister programs. There is not compensation for this part of responsibility at all (admin-4). However it does not seem to be a real burden for either of the directors, just as director G said: “It (evaluating sister programs) is totally voluntary and pressure-free. It is an experience sharing process.” Just like what the administrator told me that this evaluation is not a judgmental process, it is a way to help and support those “weaker” programs.

Grateful. Despite the high pressure as well as many responsibilities that Demonstration program directors experienced, the feeling that was shared mostly by director E was gratefulness. The assessment, for her, gave a chance or a push to work on cleaning up and reorganizing things: “It is very good! It allows and pushes programs to revisit everything so that some necessary reorganization and clean-up are identified and done” (director E).

It also provides encouragement and acknowledgement for programs to move on. “The evaluation also acknowledges programs’ effort and merits. It also provide[s] an opportunity for us to learn from each other” (director E).

Director E also felt that it was a learning opportunity. She went through a mind switching process from frustrated and nervous to confident and prepared. During this process they learned about things that they were not used to or with which they were not familiar.

When we finally learned more about what a Demonstration kindergarten looked like and also after rounds of discussions and revisiting (the materials) we felt less nervous and frustrated about it. For example, we made use of material more age-appropriately, and encouraged teachers to engage in the process more actively. Another way is to learn from experts within our district. (director E)

Furthermore, she felt motivated when she saw progress in some sister programs that she and her team helped:

Sister kindergartens (kindergartens in the same group but do not reach Demonstration level) will submit their self-appraisal report upon evaluation day. You can see great progress in Jiguan Yuan (kindergartens that are operated directly by governmental institutions, exclusively recruiting children from families in which one of the family members works in this institution. Usually viewed as a privilege to be enrolled). (director E)

This also resonates with what the administrator mentioned about the sense of achievement that Demonstration program leaders feel after progress was seen in the sister programs.

Challenges and concerns. The biggest challenge for the two program leaders was around document collection and organization. For Director F the most difficult thing was to keep up recording and collecting documents. They did not realize this need until they were asked to show evidence of what they had done in their work. Director G, similarly,

felt that the organization of documents was the most challenging part in the preparation for the assessment.

They also look at the documents, which are hard to organize. However they only spend one hour in examining the documents. I think it is very unpractical, because it is very difficult to manage all the documents, including meeting minutes, recording, summary, plan for teaching and educational research, class arrangement conference, family and program connection, big events, safety, hygiene and health, etc. (director G)

Other concerns shared by directors included the lack of time to demonstrate everything in only a half day evaluation (director E), and the high frequency of the assessment (i.e., once a year, director G). The concern for the frequency was not merely due to the high pressure that it causes each time, but also due to a lack of efficiency in the assessment:

I think that to evaluate once a year is too frequent. I think 2-3 years maybe more appropriate, because some of the places that need to be improved can only be improved in 2-3 years (one year is too short for those changes). I feel that the feedback looks the same every year. (director G)

She thought the same went for sister programs in that they received similar feedback each year. In addition, being assessed once a year did not give weaker programs time to make suggested changes: “This has something to do with the structural condition. For example, it is difficult for some street kindergarten to change their structure due to financial reasons. You should not use the same criteria on them as you use for public kindergartens” (director G).

Teacher experience.

Tired. The first feeling that came up to teachers' mind was "tired" when they were asked about how they felt when they applied for the Demonstration program rating. Since Demonstration level is the highest quality level so far, it has high as well as very specific requirements for every single domain of the program.

There was definitely pressure. For example, at the time when we applied for the assessment, many teachers, especially young teachers, did not have a life outside of school. Even after they went back to the dorm they were still working on their jobs brought back from school, such as making teaching material. (teacher N)

Purpose of the rating is to make you better. After being rated as Demonstration programs there was an obvious switch in terms of the pressure level that teachers experienced. Even though they still got nervous when an assessment was coming, they were more confident because they had normalized the high standards into their daily routines. They were more and more familiar with the standards and kept on meeting the standards everyday so they were not worried about not passing reassessment. In addition teachers trusted the system: "Because I think that the purpose of assessment or reassessment is not to make you fall, like what you might think. Rather the purpose is to make you better" (teacher N).

Teachers viewed the annual assessment as a learning opportunity for them. During this process they not only learned ways to meet standards, but also ways to distribute the time more efficiently (teacher W; teacher L). Teachers also felt that it was a push for them to keep up the good work, and an opportunity to receive support from other programs and experts.

Pride. As teachers working in Demonstration programs, they did not hide their feeling of pride from working in such a program. Like the directors, they were very much aware of their role in helping sister programs and were proud to be able to help.

It was the feeling that we all need to put effort in, so that once we reach that goal everyone [would be] very happy. The reason that I think it is a push effect is, not only on the day of assessment, that when we open our door to sisterhood kindergartens and they learn from us our teachers gradually sense that we are one of the best. So they will have strong sense of identity and self-realization.
(Teacher N)

The notion of a ‘push effect’ is a powerful illustration of a higher level program imparting and encouraging another program to move upward in their quality and simultaneously passing along self-confidence and pride in their work.

Support. Even though no financial incentive was associate with rating level, other support, especially professional development support, was available for teachers. Professional development opportunities that were provided by the government were more abundant for programs with higher rating level. This included workshops, trainings, and coaching opportunities. For instance, teachers have the opportunity to work in MOE for half a year so that they can go with the MOE administrator to different programs to do the evaluation. The administrator becomes a mentor to that teacher for that half of a year. “It is very helpful,” according to the teacher who was working in MOE at the time of the interview. It was helpful since they have a better understanding of the program quality and at the same time learn some leadership and management skills from a team of leaders in MOE.

For programs that were not at the Demonstration level they had support from both the Demonstration program and the government. When a program planned on applying for a higher quality level they let the Demonstration program of their learning community know, and then expected leaders or master teachers from the Demonstration program to come and conduct a pre-rating with the standards of the level they applied for.

What was found from program directors and teachers mostly resonated and matched with the implementation strategies described earlier. Their reactions also reflected the effectiveness of implementation. The results showed that they were satisfied with the system and bore a sense of responsibility around being an active part of the system.

Implementation Stage of KQRS

The KQRS is not like many other QRIS systems. It is both old and new; it has separate components but they are integrated in a complex way. When viewing the KQRS's implementation as a whole, it is even harder to define the stage that it is based on in the framework of implementation science. It is old because KQRS was launched in 1989, 26 years ago and it has been run under a stable, functioning system for years. It is new in that Xicheng district is now adopting a new way of working in this system and it works well so far. It is new also because it is missing several critical pieces that are viewed as important from the perspective of implementation science.

There is no doubt that leadership is in place in organizing the system and implementing practices even after the joining of the two districts. As previously described, the leader was committed to the system and was supervising both districts

using the same policy and criteria. It was also clear that this system managed to provide a hospitable environment for its development. Programs were encouraged to engage in regular communication with each other, sharing research topics and ideas.

The protocols for staff selection and training have been in place for years. Assessors are selected from experts who have a strong reputation and expertise in certain fields. Trainers for teacher professional development are promoted from schools and are also considered experts. There is no performance assessment protocol for assessors, but there is a research task for professional development trainers. They are in charge of research and studies in different domains related to professional development. This selection and training strategy has been working for years and it seems that it is not considered problematic or a barrier for the system because it works well. Directors and teachers trust the experts to conduct assessments and provide feedback and suggestions. The fact that no standardized measurement tool is used in the assessment makes it less necessary to conduct trainings on the use of instrument and almost impossible to conduct reliability checks. Thus from the perspective of implementation science the competency driver is not fully developed.

The installation of the new way of work may be viewed as a returning path that KQRS takes to update the system. The notion of Learning Community was brought up by the administrator, who takes the major leadership role in the system. It comes from her value of creating a harmonious system that includes every program and provides every program with a full range of support from the government and from parallel sister programs. In the exploration stage programs, especially Demonstration programs, were

asked how much they would be willing to take the responsibility to evaluate lower-level programs and be their mentor on a daily base. But other than this general investigation little data were collected to show the feasibility of the new system.

Despite some resistance from Demonstration programs this new system was installed and companioned with the gradual promotion of buy-in among programs. Since 2006 when the new system was installed now it has reached the goal to include all the programs in Xicheng district and set up learning communities all around the district. However other than some programs showing improvements in their overall level of quality, little evidence has been collected to show what is still missing or how the essential elements might be improved upon. Benefiting from the existing system that had a stable and functioning competency driver, organization driver, and leadership, the installation of the new system was not as hard as starting from the beginning to develop the system.

CHAPTER VI

DISCUSSION

One of the purposes of this discussion is to integrate different pieces of the results into a meaningful and holistic portrait of QRIS for North Carolina, Singapore, and Beijing (see Appendix C for cross-cultural comparisons). The second purpose is to look across the three QRISs under the three contexts, which possess different cultural, historical, and political backgrounds, in order to have a deeper understanding of the way that each is implemented now. Thirdly, the discussion will focus on implications for an implementation science framework based upon the cross-examinations of the three systems and how the framework might be adjusted to fit with the context of each country. Finally, potential future directions for each system as well as for research will be discussed at the end.

Integration of the Results

North Carolina Star Rated License System

It took North Carolina's early childhood program quality rating system 13 years to grow from an assessment-only AA system to a comprehensive system that incorporated assessment, improvement, financial incentives, and customer education. Even though some of the components are still missing, under developed, or not connected with the other components well, it does provide a relatively comprehensive picture of a quality rating and improvement system.

Among all the components, standards are the foundation that shape the whole system. In the case of North Carolina's standards, most of them are evidence-based. For example, the fact that education is weighted half of the total quality points is based on accumulative empirical studies that showed the importance of teacher and director education in affecting classroom quality (Denny, Hallam, & Homer, 2012; Mims, Scott-Little, Lower, Cassidy, & Hestenes, 2008).

Recent research evidence showed that the current rating system is not able to distinguish the very good from good quality (Hestenes et al., 2014). The measurement tools that have been used for years are also under question. According to the feedback from teachers and directors it seems that assessment is the least satisfactory component among all the components of QRIS. It is not surprising since this is a high-stakes and high-pressure assessment, and teachers are under evaluation for several hours. Compared to other components, such as feedback, financial incentives, and quality improvement activities, assessment could be the most stressful event for both teachers and directors. When asked about their perspectives toward assessors, most of the interview participants showed neutral affect and impression. Thus, it seems that it was the measurement tool or the way that it was conducted that caused the dissatisfaction toward assessment. This was confirmed through both teacher and director interviews. They found it very hard to meet some standards described in the scale items (teacher C; director B). Based on these broader considerations, a group of researchers are now funded to develop a new measurement tool which aims to depict a portrait of a program rather than focusing only on individual classrooms, and by emphasizing teacher-child interactions and learning

oriented practices that will more effectively enhance children's positive experiences in programs. It also aims to add a focus on program administration, organizational climate, and provide practical guidance for quality improvement.

This group of researchers have collaborative relationships with the agency administering the QRIS and with the quality assessment agency. Thus there is a strong tie established between research and policy. The utilization of external resources (i.e., the research institution) and the communication across the leadership teams combine to better align practice with policy. In fact since the leaders of the assessment system are also researchers, data collected in assessment are consistently used to inform their decisions. This kind of data utilization maybe achieved even if the leaders were not researchers themselves, however it seems more likely to be the case when the two roles are combined. This kind of combination can also be seen in other levels of leadership. For instance, the supervisors of assessment were assessors in the past and required to be highly familiar with the instruments. This ensures that supervisors have enough knowledge and experience to instruct fellow assessors. This combination between the research institution and the QRIS administrators turns out to be effective in turning research results into practice.

The development of the new measurement tool as well as the formation of the advisory committee drew the system back to revisit an earlier stage featured with frequent exploration and consultation with different groups. However this does not mean the whole system is being pulled back. While the system is functioning as it has been for years, this reoccurrence of stages in the system level indicates an upcoming update of the

system. For intervention programs, it could be a scale-up process. For QRIS, however, it is a process that aims for higher fidelity and more buy-in among stakeholders.

This updating is supported by a hospitable administrative environment that already exists. With staff and personnel properly selected and rigorously trained, this system is able to enact new strategies in a relatively efficient way. With a strong leadership and a developing data system, problems that emerged during the updating should be addressed in a timely manner. During the exploration stage of the updated system, communications with programs becomes more frequent. However it could be a concern if this communication between the government and programs is not maintained once the new system is installed. A communication loop does not appear to be strong in the current system. There may be occasional and personal communication between programs and personnel, such as between assessors and consultants. However it is not systematically used to inform decision-making. A lack of communication loop may result in a disconnect between policy and practice. According to implementation science, establish and nurturing a communication loop, either within the administration level or between administration and the participants is key to building a hospitable administrative environment. Thus, it is important to draw to the administrator's attention that a strong communication loop may be missing.

The updating, which will be based on the advisory committee's recommendations, may address several concerns that programs have for this system. The disconnection between assessment and improvement has been a concern for both administrators and programs. Even though some programs with relatively abundant resources have managed

to hire people to supervise and provide regular support for quality improvement, this is a luxury for most of the programs. The good thing is that programs with fewer resources are more likely to be helped when the TA resource is scarce. However, the helping is limited and scattered, although effective when it is done well. Both directors expressed their concern about creating an environment that meets the standards in the current rating system due to a lack of financial resources. In the new measure, however, less emphasis is put on setting up the environment and more is on the utilization of the environment. This could potentially reduce some of the burden from programs' shoulders and allow some extra money for them to improve other aspects, such as teacher professional development, which is another concern for programs that have fewer resources to pay or send teachers to continuing education courses or workshops.

Another possibility to increase the effectiveness and efficiency of improvement is to coordinate the staff training process for assessors, consultants, and TA providers. For now they are disconnected and show little coordination and consistency.

The lack of consumer education is not a problem for the system, but more of a problem for the whole society. With the increasing education levels for parents, customer education might be more successful than in the past. However, in order to promote buy-in, in parents and, in a broader sense, in the public simply increasing fidelity of the system may not be enough. Through the interview it appears that program leaders and staff do not invest much in parents' understanding of the importance of quality, or what quality is about. They are content with parents knowing that more stars represents better quality. Financial incentives, from another angle, actually indirectly compensate for this

lack of direct customer education by rewarding high quality programs and families who enroll in the high quality programs with financial support to enhance parents' awareness in searching for these programs intentionally. So it should be the responsibility of everyone, including the researchers, policy makers, administrators, program directors and teachers, to spread the information about the benefit of a high quality program and how to have access to these programs.

In general, North Carolina Star Rated License system demonstrates comprehensive functions in supporting programs to grow. There are flaws associated with the alignment and communications among different parties, and some inefficiency associated with the measurement tool and strategies. However, it shows great flexibility and capacity for making adjustments and changes to existing policies and practices.

Singapore Preschool Accreditation Framework (SPARK)

The development of SPARK is quite a legendary story within QRISs. Within its very short history (seven years), including the two years of measurement development and piloting, it has grown to a full system that has its own measurement tool, data system, training protocols, and committed leadership group. Leadership is the core driver of the whole system. It is also reflected in the system's philosophy as well as within the measurement scale. The administrative leaders played an essential role in the whole course of development and implementation. They initiated the visits to countries where QRIS had been implemented for years. They consulted with experts on validity and reliability issues. They also conduct piloting studies within programs. This could not be done by one or two people, however it was the leadership team that drove the whole

process from the very beginning. After five years of implementation when they faced the point where a reexamination is possible, it was the leaders who determined what to look at and consider what else might be missing.

With little coordination with external institutions, the early childhood education department in MOE functions independently and makes their own decisions. This strategy may be the key to its fast and efficient implementation. There are no external agencies or advisory groups of stakeholders that are directly involved when a decision is made; the implementation becomes immediate. However, the disadvantage is obvious too. The system may lose potential resources and ideas that could provide support to the existing leaders and programs. Other than the monthly committee meeting no obvious feedback loop could be seen. It seems that this system is strongly supported by MOE so that it does not really need support from elsewhere. Thus, leadership is the key to success in this type of system.

One of the most salient advantages of SPARK lies in its ability to promote buy-in with programs in such a short period of time. The fast start for this system caused some fear and anxiety for program directors and teachers at the beginning. The administrator was very much aware of the sense of resistance and fearfulness from programs. Thus, SPARK devoted lots of resources and energy to help programs through the process, including providing convenient and flexible support, being accessible in timely manner, and training staff to be friendly and encouraging all the time. The tenet of SPARK, at this stage, was to start programs at an equal pace, and try to avoid competition. It is hard to say how this may change after moving to a different stage, however, presently, those

strategies worked so well that both program directors and teachers expressed their gratefulness in receiving the support. However, the support is primarily at the instruction level at this stage. Programs still need to find resources, such as professional development resources and financial resources themselves. It is, again, the lack of cooperation with external institutions that narrows its' ability to provide additional support.

SPARK intentionally built a community-like system for programs, at least during the early stage of implementation, by discouraging competition and encouraging development at the same pace. The use of this strategy is also a way to increase a program's motivation to participate. Since SPARK is voluntary it is not supposed to be intimidating for programs from the very beginning. However, despite that it is voluntary it did feel compulsory for some programs. It has become an identity that programs would like to have in the early childhood education field. The fact that the certification was made visible to the public also put a type of 'invisible pressure' on programs to participate. The use of the invisible pressure and strong instructional support successfully attracted programs to apply for the rating.

The notion of a "journey" was emphasized by all the participants from Singapore in this study. This notion was invented to help programs understand the purpose of the system. The term contains a lot of information, such as it is an ongoing process, this is not a competition or test, and support is available along the road. It is actually very common for program directors to call the administrators directly for questions and concerns. It convinced programs that this system will not be a burden, rather it could help

them find ways to improve. What was reflected in one director's interview was that SPARK did help them identify problems and helped narrow the goals and direction for them. The rapport established between the government and programs turns out to be an important part of SPARK's philosophy.

Results from the survey matched the interview results to some degree. Teachers who participated in the survey thought that they had a moderate amount of knowledge about different components of SPARK, and moderate to high level of satisfaction toward the same components of SPARK. These results are consistent with the interview in that teachers were in the process of learning, and were satisfied with the process since they received lots of support. The effectiveness of SPARK received 3.86 out of 5. From the interview three out of four teachers explicitly said that it was effective. Thus it is consistent that most teachers, but not all, thought it was effective in enhancing program quality. In addition, teachers' satisfaction toward SPARK was positively associated with their perception of its effectiveness. These results indicate that teachers who were not satisfied with SPARK were less likely to think that SPARK was effective. But given the short implementation history it may need more time for its effect to be seen. Thus it is important to conduct research to examine its effects on programs as well as on child outcomes.

The developmental progression of SPARK demonstrated a live example of how a system is developed, installed, implemented, and sustained. Through the course it demonstrated high efficiency and a great ability to make adjustments to existing policies. The most important driver turns out to be leadership, which showed great power in

promoting buy-in in the installation and initial implementation stages. Moving forward to a full implementation stage while sustaining the work in the current stage is on SPARK's agenda in the near future.

Beijing Kindergarten Quality Rating System (KQRS)

Xicheng is one of the districts out of 16 districts of Beijing city. The KQRS in Xicheng, while aligning with the general guidelines of KQRS, demonstrates a unique case of KQRS implementation. It likely does not represent any other districts in China or in Beijing in general sense. However it does carry on some key features of Beijing's KQRS, such as standards, measurement, and staff selection and management.

The fact that standards and measurement directly overlap with each other provides a clear alignment between the two. However the problem is obvious too. While it is hard to tear apart the two it is hard to know what each standard looks like on each dimension of quality. For example, a standard says “conducting proper child assessment.” Experts walk in a program and make a judgment on if the assessment tools teachers' use are appropriate. There is little evidence to show why the tool is appropriate or why the way it is conducted is appropriate. Or in other words, the absence of a standardized measurement tool allows many potential opportunities for bias and subjectivity in making decisions on a program's quality. It also provides little instruction simply by looking at the rating scores for each item.

Assessors are selected based on their expertise and reputation in a certain field. There is no specific protocol or criteria for the selection. The experts are invited, rather than hired by MOE to help assess program quality. It is hard to say why this clear

absence of a standardized measurement tool and a specific recruitment protocol has been a feature of KQRS for so long. It could be that there is a lag in people's awareness in the value of using an evidence-based tool to guide behavior, so that in the early stage no one thought about using standardized tools. It could also be the philosophy and values that guide people's behavior in this country is just different from the other two countries. The initial developers and leaders may not trust quantitative methods to provide lively and rich evidence of program quality. Similarly, since the assessors are already experts in certain fields, there is no single criterion that would fit for everyone. However, all of these assumptions are only assumptions. It is clearly important to study the impact of these processes within the Xicheng context before a recommendation is put forward that this situation needs to be changed.

One of the unique aspects of the KQRS in Xicheng is its three-level rating system and the Learning Community. The advantage of this strategy is that it saved time and money since there was not a need to train people who were not from this field or had not been in a classroom and understood child development and classroom management. The disadvantage is the danger of bringing personal bias into the assessment, especially when there is neither a systematic training for evaluation nor a standardized scale to use in the evaluation. The success of this strategy in Xicheng is largely due to the strong leadership. The value that the leader holds is to build a harmonious community for programs to support each other, especially for those advanced programs to support other programs. Leadership turns out to be the key element in the process.

As for the effect of implementation, results from the survey showed that teachers are most satisfied with the professional development associated with assessment. They also perceive themselves to know the most about this part compared to other components of KQRS. Teachers were neither satisfied with nor perceived themselves as knowing the group assessment well. It is mainly due to the selection of the sample. All three programs are demonstration programs which are in charge of assessing other non-demonstration programs. However, it is only the directors and group leaders who go and conduct the assessment. Thus it is not surprising that teachers were mostly not aware of it and consequently not satisfied with it. This result indicated that the learning community did not involve everyone. Teachers, especially teachers in demonstration programs were not involved or minimally involved in this process. If the purpose of building the learning community is to give every program opportunity to learn then it should include everyone, including teachers.

Results from the interviews indicated that both directors and teachers thought positively about the system overall. The sense of community responsibility in helping sister programs improve is consistent with the administrator's value in that the more advanced programs should help and facilitate programs that are less advanced. Even though demonstration programs receive no compensation or reward for evaluating others, none of them felt it was a burden. This may be partly due to the expectations from the whole society, and partly due to the sense of achievement they acquired upon seeing others improve. The annual assessment was thought of as a powerful way to maintain program quality when they first thought about it. The administrator was also proud of this

strategy since it is not common in other districts. However, according to the director interviews, they felt it was not necessary to have an annual evaluation since they found that the feedback each year was similar and that one year was too short for suggested improvements to happen. It is clear that this thought was not communicated to the administration level yet. Strengthening the feedback loop between government and programs might be an important next step in this system.

After decades of implementation, KQRS still needs some adaptations and new strategies to evolve into a better system. The absence of a standardized measurement tool may be adaptive in the cultural and political context of China, however there is a disadvantage in being able to compare and predict future directions. It is impossible to use hundreds and thousands of qualitative data points for those purposes. The learning community saved money and energy for the government and also promoted strong collaborative relationships among programs. However, the government as well as programs should be cautious about the bias and subjectivity that this approach may bring into the community evaluation since this may have a significant influence on each program's development.

Contextual Comparisons across the Three Systems

The three systems are similar in many facets, but also differ on several dimensions. The purpose of this section is to have a better understanding of QRIS by looking across the similarities and differences of the three systems in three different contexts. It may also shed light on the implications for implementation science and its application in different contexts, which will be the discussion in the following section. In

the current section, rather than following the structure of implementation science as it was done in the result section, I will use more general concepts as the themes. By doing this I am more likely to capture themes that are not yet in an implementation science framework. Thus the themes may be within or beyond an implementation science framework. The themes were developed because they showed important roles in the implementation in one or more systems (see Table 5). The themes are: Leadership, organizational climate and communication style, and performance assessment. The discussion will focus on the comparison of each theme across the three regions and then try to understand them under specific cultural, political, and historical background of each country.

Table 5

Discussion Themes

Leadership	
North Carolina	Facilitative leader; Form coalition and partner relationships with leaders in other domains.
Singapore	Directory and participative leadership combined
Beijing	Directory leadership with some decentralized responsibilities to senior staff (i.e., demonstration program directors)
Organizational climate and communication style	
North Carolina	Business-like relationship; Straightforward; Task-oriented; Cooperation is largely seen; Formal
Singapore	Face-to-face; Rapport between government and programs; Independent agency; Balance between informal and formal
Beijing	Close relationship among programs; Mixed relationships as friends, family, but also colleagues, and superior-subordinate; Frequent communication; Lots of informal communication

Table 5

(Cont.)

Performance Assessment	
North Carolina	Used for staff selection and training; Look for evidence of the effectiveness of the system, including measurement validity and fidelity of the whole rating system in differentiate quality levels
Singapore	Three pilot studies with programs before launch; Reliability check in the first round of staff training; Reexamine measurement validity after five years' implementation
Beijing	No formal data base

Theme 1: Leadership

NC star rated license system. Leadership is one of the drivers in implementation science. It can be seen in all three systems. Leadership that is discussed here includes adaptive leadership and technical leadership, as well as incorporates the beliefs, values, and resources that support the leadership. There is no doubt that leadership is very important in successful implementation. However, variations exist across countries in the role that leadership plays and the support it receives in the system. As previously described, North Carolina is led by DCDEE, for which the mission is described as “in collaboration with our partners, to protect the health and safety of all North Carolinians and provide essential human services” (DCDEE website). It is very clear in this mission statement that collaboration with partners is the key in organizing and management. This leadership strategy is reflected in practice as well. In 2009, the formation of the advisory committee by stakeholders from across the state demonstrated the leadership that is built on collaboration and partnership. Throughout the interviews there are seldom “I”

statements, instead most of the time it was the “we” statement that occupied both administrators’ narratives. It indicated that any decision was not made by any single person, but was made collaboratively with others. The leaders are considered facilitators or organizers rather than commanders or decision-makers in this type of system. It seems that this is the trend and data guide the direction of the system. Leaders’ values play a less important role here.

SPARK. In Singapore there were many more “I” statements than was in North Carolina’s administrator interview. Through the exploration stage it is clear that the administrator took the main responsibility for consulting with experts all over the world, making decisions on the structure of the scale, the values statement, and conducting pilot studies. However, the “I” statements did not dominate the whole conversation. The quality assurance department is just one part of the system, although it supervises the whole system. In the development phase other departments also engaged in the process. Leaders have the last say on the issues, however it is based on committee opinions and data. This strategy is also reflected in the process of generating assessment reports. As described in the result section, the generation of the report goes through several thresholds before it is finalized. The last threshold is the administrator who reads through the previous version, raises questions, and finalizes it. Leaders in SPARK are expected to make final decisions, however they also actively participate in discussions about the decision and adjust their decisions based on a more collaborative approach. Leaders’ values play a moderately important role in leading the system. Leaders are decision-makers and also participants.

KQRS. In Beijing, “T” statements were used more frequently than in the other two countries. The entire idea of building a learning community was based on the administrator’s values and beliefs about including every program in the system. The leader was expected to provide the direction for the system’s development. Thus, the leader should be clear about the direction and have the ability to guide the system toward that direction. The administrator demonstrated great pride and rapid action in making decisions. In fact, decisions were not made immediately after the idea came along. There were investigations with programs about the feasibility of the new way of work before the act was issued. Even though some demonstration program directors were against this proposal it did not really change the initial idea. In 2006 this new act was issued with a few programs joining and trying out this new way of work. Later on, more and more programs joined willingly because they saw the positive effect of the learning community on the programs that experienced it. This strong leadership style is efficient and effective so far. Leader’s values, beliefs, and personal styles play a critical role in this kind of system. Leaders in this system are considered as the absolute decision-maker and direction-guider.

Context Comparison. As described earlier, the leadership style in North Carolina is based on partnership and coalitions among different groups of stakeholders. In Singapore the leaders are mostly decision-makers but also active participants in decision-making activities and show a certain level of democracy. In Beijing, leaders are expected to find and lead programs to a direction and also make decisions. The sociocultural values and beliefs shape the leadership style and also its effect on the whole system.

Americans, in general, hold beliefs like liberty, independence, individualization, and democracy (McKay, 2013). The practices reflect the essential ideology of the U.S, which is the freedom in faith and democracy (Gray, 2012), and also the political strategies based on its diverse population composition (Gray, 2012). They tend not to believe in authority without questioning and testing. In America class is not as important as it is in other cultures. According to a chart that shows the distribution of 50 countries on a hierarchical/egalitarianism scale, U.S. falls on the closer end to egalitarianism, right next to the extreme end of egalitarianism. On the same chart, both China and Singapore fall on the range that is closer to the hierarchy end, right next to the extreme end of hierarchy (Solomon & Schell, 2009). Although this chart has the potential to over-generalize or simplify large heterogeneous groups, it does provide the basis for us to interpret the difference and similarities across countries.

Some countries are strongly influenced by Confucianism, which emphasizes the class and hierarchical rankings in society, and also values loyalty and obedience to seniors and authorities, such as teachers, parents, and leaders, who are very much valued and respected. When thinking about the leadership style in this study, it is not hard to understand why in China people look up to the leader, rely on him/her to make decisions, and also respect and follow the leading without much questioning and resistance. Since Singapore is composed of 70% Chinese, who still carry the traditional ancient values about ranking and hierarchy, some of the same practices are observed in their leadership style, where leaders are expected to make most of the decisions and lead the team.

Singapore is small and relatively young country that relies on fast growth in its economy to support its acquisition of resources. Thus, an efficient government is needed (Siong, & Chen, 2007). Too much democracy may slow down the process. However it is equally, if not more, important to be pragmatic. By saying pragmatic, this means that all the actions are result-oriented. Seldom can anyone afford to move slowly in making changes when there are limited resources. Thus, a combination of directory and participatory leadership style is seen in Singapore. This may also differ by the leader's ethnicity; Chinese are more goal-oriented, and pragmatic than their counterparts from other ethnic groups, such as Malay and Indian (Dimmock & Walker, 2005). The leader, who is a Chinese, showed a strong and very efficient leadership style in the process. This leadership style matches the nation's requirement of being efficient and pragmatic.

The fact that ECE in China is a part of the education system, and is supervised by the national education department makes it more likely to be influenced by the overall political strategy regarding education. However, while the Chinese government still carries on the leadership style from previous ages, which is control over things, the emerging consciousness about rights and citizenship are awakening (Fewsmith, 2010). The implication of this change for the implementation of a system is that leaders may begin to move toward a more participatory leadership style, while still maintaining the authoritarian role by only involving senior staff in decision-making processes, and using the senior staff as an intermediary between the leadership group and the grassroots level (Dimmock & Walker, 2005). In Beijing's case, directors from the demonstration programs are considered senior staff who are endowed with power to supervise lower-

level programs. This strategy showed quite a few advantages as described previously, such as taking off some burden from the government, increasing the connections among programs, and steadily improving program quality. Under the current historical and political environment as well as the influence of the embedded cultural values it is impossible to copy what it is done in other countries, especially with what is done in the U.S. However there may be space for improvement.

When looking at the relationship between different types of leadership styles and the outcomes reflected in director and teacher interviews it is interesting that even with a high level of democracy and liberty in North Carolina, program directors and teachers still showed different levels of concerns and complaints. In Singapore, programs showed high levels of satisfaction under the combination of dedicated and participative leadership style. In China, with the moderate to high level of control and low level democracy there also were mixed feelings toward the system. The difference may due to the difference of the effectiveness of the system, it may also be due to the difference in the communication styles as well as organizational climate of each system under the certain contexts.

Theme 2: Organizational Climate and Communication Style

NC star rated license system. In North Carolina communication between the government and programs is formal, mostly through the assessment feedback and consultants' visits. Staff (i.e., assessors, consultants, TA providers) are well trained to provide professional feedback and assistance to programs. Whenever there are questions or concerns it will be taken seriously in decision-making. For example, programs were called to investigate how difficult it was for them to put a certain requirement into action,

their opinions will be used to make adjustments in the strategies recommended by advisory committee. The relationship between programs and government agencies is like a customer and distributor. It is a business-like relationship with high professionalism, equal rights in bargaining, and clear goals and approaches in achieving those goals. Communication among agencies, however, is not as frequent as it should be. It is clear that they are connected and do show some collaborative relationships. These communication strategies are effective in that they precisely convey information and messages and it is readily used to inform practices. However, it may take longer for feedback to be received and used in practice on both sides since it needs to go through several formal steps to reach different levels and agencies.

SPARK. The communication in SPARK is mostly face-to-face. In this initial implementation stage programs needed lots of instructions and assistance to adapt to this new system. Assessment feedback is also delivered by assessors personally and they walk through the results with the program director and teachers in a timely manner. It is also easy for programs to reach out for help by making phone calls to the immediate supervisor in the district or even the head of the early childhood education department. Since there is little cooperation with external agencies or institutions, the communication at administration levels have fewer barriers and are mostly efficient. The communication with other stakeholders is less frequent and only involves feedback on assessment reports. This suggests that SPARK is very independently operated, and has its own feedback loop established with programs and within several departments in MOE. This strategy accelerated the whole process of installation and initial implementation. However, it is

hard to tell if the same small-scale communication and cooperation style will work in the long run.

KQRS. In Beijing communication among programs are abundant, especially for programs that are within one learning community. Programs within one learning community are close to each other in their physical locations. They are encouraged to communicate with and learn from each other. When a non-demonstration program has problems and questions they can easily reach out to the demonstration program of their community for help. Demonstration programs are also able to customize suggestions and plans for their sister programs within one community. The communication between different demonstration programs is much less than community-based communications. However the relationships among all the programs in Xicheng district seem to be closer than in other districts. Programs benefit from this close relationship and frequent communication among them. The feedback loops are more efficient than they used to be, and this results in a faster enhancement in the quality. Administrators also pay visits to programs frequently. The communication style among all parties are mostly face-to-face and involve strong personal relationships. Different from Singapore, the notion of relationship is also embedded in Chinese culture. Interdependence and cooperation between agencies are common, sometimes the relationship are too intertwined to separate (Gold, Guthrie, & Wank, 2002). For example the relationship between the government and program is like friends, family, but also colleagues, and superior-subordinate. It is important to consider how this relationship might affect communication loops between

them. In addition, it is hard to document all the conversations and translate this into useable data.

Context comparison. Organizational climate refers to the way the system is organized and the connections a system makes with an external environment. It may reflect leaders' personal beliefs and values to some degree (Siong, & Chen, 2007), relating a lot to leadership style that was talked about earlier. Communication style is a critical component of organizational climate. The way that people communicate with others determines the quality, efficiency, and also the interpretations of messages, which will in turn help define the organization climate of a system (Elving, 2005; Guzley, 1992). It reflects hierarchical beliefs, importance of relationships, and attitudes toward conflict (Solomon & Schell, 2009).

There can be many dimensions of communication style. However when simplified it may be viewed as varying on a range between direct to indirect. Within a direct communication style people tend to talk more about issues rather than people, use more precise wording rather than more subtle wording, and relate less to the context rather than incorporating the details of the context. Indirect communication style demonstrates the opposite characteristics of direct communication style (Solomon & Schell, 2009). The same 50 countries mentioned in theme 1 were rated on the dominant communication style in their culture on a scale. The U.S. falls in the range next to the extreme end of direct communication style. Both Singapore and China fall in the range next to the extreme end of indirect communication style. This seemingly over-generalized categorization was actually reflected in reality in the three countries in this study.

The direct communication style in the NC QRIS enables the transmission of information precisely, professionally, and effectively. Because of the devalued notion of hierarchy in this society people spend less time in regulating their behavior based on their position and ranking in the society. The goal is to focus on task achievement instead of maintaining relationships (Dimmock & Walker, 2005). This system also allows conflict to happen and since people communicate in a direct way and conflicts are simply viewed as a barrier during the course of implementation, and need to be solved through everyone's effort. An example of this is the grievance process used by the assessment agency. Programs are allowed to file a grievance against a score and then a review system is put in place to resolve the disagreement. It is viewed and handled in an objective way. As aforementioned, coalition and partnership is the key of the system. Thus connections were established with the external environment as well as within the system. The organizational climate is featured as professional and business-like. One of the disadvantages associated with the lack of personal, long-term interaction and relationship maybe the losing of information that can only be acquired through a long-term relationship. For example, what are the reasons that a program stays at a certain level and never grows or increases in star levels? There may be some visible and obvious reasons, but the subtle reasons are hard to get to unless the long-term relationship is established.

Singapore, as placed at the indirect communication end, did show many relational communication styles, such as building and maintaining the welcoming and friendly communication style with programs to encourage them to ask questions and get support. These communications, even though very friendly, still involved hierarchical levels. The

system, even though voluntary, is viewed as mandatory by programs because it is initiated and organized by the government. Since the relationship between government and program is a critical part of the cultural context, conflicts are avoided intentionally. The leaders are trying to build a harmonious environment. However, honesty and integrity is also highly valued. The government and programs are very clear about their responsibilities and obligations. They establish the relatively close relationship while still maintaining some independence, just as it is of the system itself. SPARK is independent from external institutions, such as research institutions, professional development agencies, or agencies with funding resources.

In China, the notion of community was formed long ago. Most Chinese find fulfillment in meeting communal goals (Whitehead & Whitehead, 1978). In this group-oriented society relationship sometimes overrides the task itself (Gold et al., 2002). Cultivating and maintaining a harmonious environment is valued in this system. Because harmony is likely to be damaged when conflict emerges, strategies may be taken to avoid conflicts. Similar to Singapore, despite the harmonious environment, hierarchy still exists in communication styles. This hierarchy turns out to be useful in helping program understand their responsibilities and obligations. However it also creates barriers for authentic communication. The strong harmonious organization climate also is like a double-edged sword that on the one hand builds the long-term relationship which is beneficial for long-term improvement, and on the other hand, increases the interdependence, and blurs the boundaries between different parties, which may not be good for the development of an efficient and professional system.

Theme 3: Performance Assessment

NC Star Rate License system. Performance assessment includes the assessment of staff (i.e., quality assessors, consultants, and TA providers), the measurement tool, and the assessment of the fidelity of the whole system. It is clear that the QRIS in North Carolina is evidence-based. Everything is monitored by collecting data and documenting information to prove its effectiveness. For example, assessment reliability is monitored constantly, empirical studies were conducted to provide evidence for the validity of measurement tool, and child outcome data were collected to examine the effectiveness of the whole system. Results from those assessments were further used to inform future change in the system.

SPARK. There are also ongoing performance assessments in SPARK. Validity and reliability tests were conducted for the new measurement when it was developed. After five years of implementation there is going to be another validity and reliability test with a larger sample, however this has not yet begun. The selection of assessors are based on their reliability scores. Their performance, however, is not assessed regularly afterwards, however they attend one standardized training every year. Currently, there is no attempt to examine the effectiveness of the whole system using child outcomes. This may due to the early stage of development of the SPARK system.

KQRS. Performance assessment is largely missing in KQRS, both in the selection and training of staff, and in the measurement tool. Statistics show the number of programs in different quality levels, however none of the data is used to link with child outcomes.

Context comparison. The assessment process in the three regions also follow different approaches that are connected to the contextual culture of each region. In North Carolina the assessment process is highly routinized and follows strict processes for establishing and maintaining reliability and utilizes instruments that have high levels of psychometric testing behind them. The lack of regular reliability check in SPARK is mainly due to a lack of manpower in Singapore, given that the size and the population of Singapore. The absence of performance assessment in Beijing also carries on its cultural beliefs in authorities. Since authority figures (i.e., assessors) are trusted to be absolutely right the needs for performance assessment becomes less important.

Feasibility of Implementation Science in the Three Countries

Based on the previous discussion, it is clear that the three countries are sharing some similarities, especially between China and Singapore, and then between Singapore and the U.S. However, differences are also clear, especially between China and the U.S. There is also a trend that some of the values in China are being influenced by more western values, such as the notion of citizenship, rights, and liberty. With these change the practices are also correspondingly changing. However, the cultural, historical, and political environment in each country limits the globalization process, which means one model will likely not fit with all countries or systems. In this section, I will focus on the fit of implementation science in each region, and also the implications for the original implementation science framework.

As long as the values behind implementation science are in accordance with the values behind a system then we should say that it fits with the context of that system. The

lag in action and practice does not reduce the level of fit, it simply represents a developmental stage in implementation. Thus, to examine the fit of implementation science one needs to look for the match between the values behind implementation science and the values of the culture, including the general culture of the country and also organizational culture.

North Carolina Star Rated License System

Since the notion of implementation science as well as its framework was brought up in the U.S, it is assumed that it should fit with North Carolina's Star Rated License system the best among all three systems. In fact, one of the values behind implementation science is to have three equally important drivers that promote the successful implementation. In an ideal system when the three drivers are fully functioning they should be working collectively in implementation. In reality sometimes compensation among drivers can be observed. For example, when staff selection is not ideal, staff training, and a hospitable administrative environment may compensate for the shortage in staff selection. The value is reflected in the philosophy about collaboration and partnership stated by DCDEE. The whole system is driven by the three drivers collectively rather than being driven primarily by one of the drivers.

Another important value behind implementation science is that decision should be evidence-based, including decisions about staff selection and training, ways to address administrative barriers, and performance assessment. This value is also held by NC QRIS in its decision making about staff selection, training, performance assessment, and

fidelity examination. Thus the assumption about the fit of implementation science with NC QRIS appears to stand.

SPARK

Singapore embraces a combined value of Confucianism and pragmatism. The leader takes great responsibility in guiding the whole system, while also trying to balance a direct leadership style and a more democratic leadership style. The leadership driver is weighed more heavily in Singapore's situation. There is also an absence of system intervention, which is the connection with external institutions and resources. This strategy does not appear to be a part of their system. Singapore, even though it is governed by a single party, "the government is tied at the Cabinet level, not at the public service level" (Siong, & Chen, 2007, p182). There is great autonomy for individual agencies. This independence is embedded in Singapore's political environment and has been effective. Thus, it could be that system intervention is not as important under Singapore's political and cultural environment as it is in the U.S. The influence from Confucianism makes harmonious environment very important in this society. The purpose of facilitative administration is to create an environment for better performance both on the administrative end and on the program's end. Thus, facilitative administration may be weighed more heavily in Singapore and it may also compensate for the function of systems intervention.

As for the evidence-based principle that is conveyed by implementation science, SPARK did make efforts to look at evidence for its decision-making. For instance, conducting pilot studies with programs using the new scale, examining the validity and

reliability of the scale, and reexamining the scale's fidelity after five years' implementation, indicates that evidence-based practice is what SPARK values and will be used in its implementation. In SPARK's case, implementation science framework fits well with only slight exceptions as mentioned earlier. Thus, when using implementation science as a guidance for further implementation the original framework needs to be adjusted on its leadership and organization driver. To be specific, for this framework to be feasible to use in Singapore, leadership should be emphasized as the key to the final product. Another adjustment is to strengthen facilitative administration in organization drivers, and weaken the role of system intervention, or even drop it. The effect of facilitative administration is supposed to compensate for the function of system intervention. That is, the system itself should be strong enough to provide all kinds of support that is comparable to the support one may get from multiple external resources.

KQRS

Among the three systems, KQRS is the one that showed most divergent characteristics from the framework's tenets. First of all, similar to Singapore, there is an unbalanced weighting on each driver. The leadership driver is weighted the heaviest among the three drivers. Even though staff management is in place, staff training is not. This is, as aforementioned, due to the use of experts as assessors and consultants. Since authority figures are valued in Chinese culture, experts' evaluation and suggestions are respected and valued too. Thus, the use of experts is based on the cultural values embedded in Chinese culture.

Harmony is extremely valued in building the administrative environment. Just like it is in Singapore, a harmonious environment may produce the most desirable outcome since everyone in it prefers this environment and are more likely to carry out good work. Thus, harmony should be considered as a critical characteristic in the organization driver.

In terms of the value of evidence-based practice, other than requiring documents from programs for evaluation purpose little evidence is seen in how data or other kinds of evidence is used to inform decision in, such as staff selection, performance assessment, or the effectiveness of the whole system. Decisions are made mostly based on leaders' beliefs, values, and experiences. Empirical evidence is not as valued as it is in the two other countries.

. . . we must understand the Chinese concept of truth. First of all, truth is not understood as something revealed from above or as an abstract principle, however logically consistent, but as a discoverable and demonstrable principle in human affairs. In other words, the real test of truth is in human history. (Moore & Morris, 1967, p. 12)

It is rooted in Chinese culture to rely on experience in history. Beliefs and values are formed based on one's past experiences.

Based on this difference it may be necessary to incorporate this value into the framework, such that leaders' value and experience should be weighted heavier, so that a rigid selection and training of leaders may be included as a prerequisite condition for this framework to be effective. While the high regard for authorities is embedded in Chinese culture the competency driver becomes extremely important as well. The system is

composed and run by people, rather than data. Thus, the selection of assessors should be the key. The relationship between assessors and programs could be an additional component in the competency driver since relationship is highly valued and has a significant influence on many things. Similar to Singapore, facilitative administration should be strengthened and emphasized. Facilitative administration should also be able to compensate system intervention in the support it could give.

Implications for the Original Framework

When using implementation science as a framework to guide practice it is important to keep in mind the context in which this practice is going to happen. Even in a society that holds the same or similar beliefs and values that the U.S. holds, it is important to examine the components of each driver to see how they might integrate and compensate for each other.

Future Directions

North Carolina Star Rated License System

At the full implementation stage, the QRIS in North Carolina should be working on sustaining full implementation by making sure that everything is in place and fidelity is high. It is not a simple task to maintain the performance. It may mean going back to an earlier stage if the fidelity is not as desirable as it should be. This may involve several different domains, such as the improvement in the measurement, which is already being considered in the QRIS. From the implementation science perspective, the connection and consistency among different agencies is relatively weak. For example, the training protocol for quality assessors, consultants, and TA providers are quite different. This has

resulted in inconsistent instruction and assistance that programs may receive from different groups. Thus to align the different components in a consistent way maybe the next step for the NC QRIS.

Singapore Preschool Accreditation Framework

SPARK is in its early stage of implementation however it is moving at a fast pace. Because of this, SPARK needs really to examine its fidelity frequently since it is hard to monitor the quality of the system when it is moving fast. Although the shortage in manpower was a problem for the system, SPARK has been creative in hiring retired principals and using standardized training to replace reliability checks. However, as the system moves toward a more advanced stage where high fidelity is required, it may turn out that a standardized training is not enough to guarantee the high reliability of assessment. Thus, it may be necessary to examine the validity of the standardized training occasionally to make sure that it functions in a way that is expected. Since external connection and cooperation is not encouraged under Singapore's context, it is important to make sure that the administrative environment within the system is strong enough to provide the system with necessary support, such as financial, technical, and research support.

Beijing Kindergarten Quality Rating System

The learning community that is featured in the KQRS in Xicheng district did strengthen harmonious organizational climate in this system. Since harmony is highly valued in Chinese culture it is important to maintain it. The decentralization of the power to senior staff is also a strategy to keep the system balance. Despite this effort authority is

still important in Chinese culture. Thus, it makes the selection, training, and monitoring of leaders extremely important. To facilitate decision making, it is urgent to establish a data system that could be used to monitor the implementation regularly. However, the nature of the system as relationship based and its lack of standardized measurement tool, makes it impossible to keep track of the implementation quantitatively, since relationships are hard to measure. Thus, the first thing may be to standardize the measurement tool for program quality. This change will not conflict with the value of a harmonious society, rather it will support leadership and decision making for a better implementation fidelity.

Contributions and Limitations

Contribution

The quality of early childhood education programs is important, yet it is different from elementary school, or middle school quality that are mostly based on their educational quality. The quality of ECE programs is unique due to the unique developmental stages of children that ECE programs serve. Thus it is important to have a unique protocol and implementation process in evaluating their quality. QRIS emerged in the three countries at different times and demonstrated different characteristics associated with their implementation stage as well as the broader sociocultural environment in which they are embedded.

This study is a cross-cultural multiple case study which compares three cases based on the examination of each case. This study explored the QRIS system in three different countries. North Carolina and Beijing are both one of the most advanced states

or cities in QRIS implementation in the U.S. and China respectively. Grounded in an implementation science framework the mechanism of implementation under the unique context of each country was examined. To the researcher's knowledge it is the first of its kind that uses implementation science as the frame to look at the implementation of QRIS. It is also the first study that compares the QRIS in the U.S, Singapore, and China. The three countries represent three different sociocultural environments. What was found in this study is intriguing in that the three systems look similar but also differ greatly from each other. It enables us to see the influence of the larger cultural environment on the implementation of a system. Also guided by implementation science, the three systems showed advantages and disadvantages in the implementation in different domains and to different degrees. This study also provides recommendations for future steps for each system taking both their developmental stage as well as the context into consideration.

For QRIS as a field, this study provides more examples of QRISs in different countries. It increases our knowledge of QRISs and acknowledges the diverse forms of QRIS, which in turn has implications for QRIS implementation as a whole. Theoretically, this study also examined the feasibility of implementation science in different cultures. It may not fit worldwide as its original form, yet provides a basis for adjustment in cultures other than American culture.

Limitations

Limitations of this study are primarily associated with measurement. First, this study is primarily relying on interviews, especially in the case of Singapore and Beijing, due to the difficulty in acquiring governmental documents or written procedures

associated with the QRIS in the two places. One of the characteristics of a case study approach is to have abundant information from multiple resources. Even though this study utilized data from interviews, documents, and surveys, it is clear that most of the information came from the interview and survey, which makes the triangulation among interview, survey, and document difficult. Thus, it is hard to know if some information acquired from interviews could be confirmed through documents.

Additionally, the interview and survey were designed by the researcher based on the implementation science framework. Questions were designed based more on a US model and therefore some of the questions may not have been as appropriate for other countries. The researcher also had a preconceived notion of what a QRIS should function like and thus this was the bias from which she started. Open-ended questions should have been more frequently used in interviews to allow more authentic and rich information. This is truer in a cross-culture study where there is more variability in the background of the person being interviewed.

Finally, the sample was not balanced across the three countries and therefore more information was gathered in the U.S. than in Singapore and China. For example, two administrators were interviewed in North Carolina. One of them had the big picture of the QRIS in North Carolina, while the other administrator had specific information about quality assessment. In both Singapore and China only one administrator who was in charge of supervising the whole system was interviewed. Even though this is mainly due to the absence of different leadership roles in both Singapore and Beijing, it still made the data from Singapore and China less rich than it was in North Carolina. In

addition, this case study approach targeted high quality sites in each region in order to gain the most knowledge about the systems, but it is likely that programs with fewer resources, education levels, and/or experiences may have very different experiences within each system.

Future Directions for Research in QRIS

This study was an exploratory study. It serves to inform the future direction of studies in the QRIS field. In future studies it will be helpful to probe reasons behind the facts that have been presented in this study. For example, why was there not a data system in Beijing, how the standardized training works in Singapore to replace reliability checks, and to what degree is assessment and improvement related in North Carolina?

From a cross-culture perspective in the future studies, it is important to build upon the measurement used in this study and adapt it for use in different countries or states. It may promote the understanding of what factors in implementation science are the key elements that enable successful implementation using different methods, including quantitative and qualitative methods.

Conclusions

Despite the limitations, the current study showed that each QRIS exists in a different form in different countries. Further analysis showed that there are cultural, historical, and political reasons behind the differences. Thus there is no ‘one model fits all’ reality in a QRIS. One successful model in a certain country should not be copied to another country. Similarly, strategies that work in a certain country may not work in another country. However, using implementation science as the framework to guide the

understanding of QRIS implementation in the three countries did make the cross-culture comparison possible. It made it clear that some characteristics of a certain QRIS are embedded in that culture, and are therefore hard to change. However, there are still some other elements in the system that could be improved upon based on its current status. The end result is to benefit children and their families no matter what the cultural background might be and it is clear that in each of these countries this ultimate goal is at the heart of their systems.

REFERENCES

- Adams, J. Q., & Strother-Adams, P. (2001). *Dealing with diversity*. Chicago, IL: Kendall/Hunt Publishing Company.
- Ang, L. (2012). *Vital voices for vital years*. An Independent Study Sponsored by Lien Foundation.
- Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Development Science, 12*, 140–153.
- Cassidy, D. J., Hestenes, L. L., Hansen, J. K., Hegde, A., Shim, J., & Hestenes, S. (2005). Revisiting the two faces of child care quality: Structure and process. *Early Education and Development, 16*(4), 505–520.
- Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly, 20*, 345–360.
- Census Bureau of China. (2010). *National data*. Retrieved from <http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0301&sj=2010>

- Choo, K. K. (2010). The shaping of childcare and preschool education in Singapore: From separatism to collaboration. *International Journal of Child Care and Education Policy*, 4, 23–34.
- Clifford, R., Reszka, S., & Rossbach, H. (2010). *Reliability and validity of the early childhood environment rating scale*. Chapel Hill, NC: FPG Child Development Institute, University of North Carolina.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26, 673–682.
- Colwell, M., & Lindsey, E. (2003). Teacher-child interactions and preschool children's perceptions of self and peers. *Early Child Development and Care*, 173, 249–258.
- Connor, C., Son, S.-H., Hindman, A. H., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, 43, 343–375.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, 4, 50–65.

- Denny, J., Hallam, R., & Homer, K. (2012). A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education & Development, 23*, 678–696.
- Dimmock, C., & Walker, A. (2005). *Educational leadership: Culture and diversity*. Thousand Oaks, CA: Sage.
- Downer, J. (2013). Applying lessons learned from evaluations of model early care and education programs to preparation for effective implementation at scale. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 157–170). Baltimore, MD: Paul H. Brookes.
- Early Childhood Development Agency. (2014). Singapore Pre-school Accreditation Framework Quality Rating Scale.
- Elving, W. J. (2005). The role of communication in organizational change. *Corporate Communications: An International Journal, 10*, 129–138.
- Fewsmith, J. (2010). *China today, China tomorrow: Domestic politics, economy, and society*. Lanham, MD: Rowman & Littlefield.
- Fixsen, D., Naoom, S. F., Blase, D. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute (FMHI Publication# 231). CIT0009.

- Fuller, B., Holloway, S. D., Bozzi, L., Burr, E., Cohen, N., & Suzuki, S. (2003). Explaining local variability in child care quality: State funding and regulation in California. *Early Education and Development, 14*, 47–66.
- Gray, V. (2012). The socioeconomic and political context of states. In V. Gray, R. L. Hanson, & T. Kousser (Eds.), *Politics in the American states: A comparative analysis* (pp. 1–30). Thousand Oaks, CA: Sage.
- Gold, T., Guthrie, D., & Wank, D. L. (Eds.). (2002). *Social connections in China: Institutions, culture, and the changing nature of Guanxi*. West Nyack, NY: Cambridge University Press. Retrieved from <http://www.ebrary.com>
- Gormley, W. T., & Lucas, J. K. (2000). *Money, accreditation, and child care quality*. Working paper series. Retrieved from <http://files.eric.ed.gov/fulltext/ED446851.pdf>
- Guzley, R. M. (1992). Organizational climate and communication climate: Predictors of commitment to the organization. *Management Communication Quarterly, 5*, 379–402.
- Halle, T., Metz, A., & Martinez-Beck, I. (Eds.). (2013). *Applying implementation science in early childhood programs and systems* (pp. 21–42). Baltimore, MD: Paul H. Brookes.
- Harms, T. (2013). *School-Age Care Environment Rating Scale:(SACERS)*. Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early Childhood Environment Rating Scale, Revised Edition*. New York, NY: Teachers College Press.

- Harms, T., Cryer, D., & Clifford, R. M. (2003). Infant/toddler environment rating scale-revised. *Infant and toddler environment rating scale revised*.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale Revised Edition (FCCERS-R)*. New York, NY: Teachers College Press.
- Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., . . . Cassidy, D. J. (2014). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. *Early Childhood Research Quarterly*, 30(Part B), 199–214. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0885200614000672>
- Honig, A. S., & Lim, S. E. A. (1998). Singapore childcare and early education. *Early Child Development and Care*, 144, 1–4.
- Hu, B., & Li, K. (2012). The Quality Rating System of Chinese preschool education: Prospects and challenges. *Childhood Education*, 88, 14–22.
- Hu, B., & Szente, J. (2009). Exploring the quality of early childhood education in China: Implications for early childhood teacher education. *Journal of Early Childhood Teacher Education*, 30, 247–262.
- Kirby, G., Caronongan, P., Malone, L. M., & Boller, K. (2014). What do quality rating levels mean? Examining the implementation of QRIS ratings to inform validation. *Early Childhood Research Quarterly*, 30(Part B), 291–305. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0885200614000878>

- Li, K., Hu, B. Y., Pan, Y., Qin, J., & Fan, X. (2014). Chinese Early Childhood Environment Rating Scale (trial) (CECERS): A validity study. *Early Childhood Research Quarterly, 29*, 268–282.
- Lim, S. E. A. (1998). Preschool in Singapore—A historical overview. *Early Child Development and Care, 144*, 5–12.
- Lindseth, A., & Norberg, A. (2004). A phenomenological hermeneutical method for researching lived experience. *Scandinavian Journal of Caring Sciences, 18*, 145–153.
- Liu, Z. (2011). Chinese kindergarten education quality evaluation: Study based on kindergarten quality investigation in eleven provinces. *Educational Science Publishing House*.
- Liu, Z. (2012). Current status of kindergarten education quality in China: Compare to that in 1992. *Studies in Preschool Education, 2*, 3–10.
- Liu, Z., Liao, Y., Chen, Q., Yi, L., Zhou., J., Wang, H., . . . & Zhou, X. (2011). *Early childhood education quality evaluation*. Beijing: Educational Science Publishing House.
- Liu, Y., & Liu, F. (2007). The study of the impact of teacher-child ratio on preschool teacher's teaching reflection. *Teacher Education Research, 6*, 62–67.
- Liu, Y., & Pan, Y. (2008). Development and validation of Kindergarten Environment Rating Scale. *International Journal of Early Years Education, 16*, 101–114.
- Liu, Y., & Shi, J. (2011). The status and concerns of early childhood education after the publish of National Ten. *Early Childhood Education, 24*, 1–6.

- Lloyd, C. M., Supplee, L. H., & Mattera, S. K. (2013). An eye to efficient and effective fidelity measurement for both research and practice. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 139–156). Baltimore, MD: Paul H. Brookes.
- Lower, J. K., & Cassidy, D. J. (2007). Child care work environments: The relationship with learning environments. *Journal of Research in Childhood Education*, 22(2), 189–204.
- Lu, Z. (2014). *Preschool teachers, children's and their parents' perspective on perfect kindergarten in Shanghai* (Master Thesis, East China Normal University).
- Luo, J. (2013). Governmental responsibility in early childhood education in China (Master's Thesis, Minzu University of China).
- Ma, X., Shen, J., Kavanaugh, A., Lu, X., Brandi, K., Goodman, J., . . . Watson, G. (2011). Effects of quality improvement system for child care centers. *Journal of Research in Childhood Education*, 25, 399–414.
doi:10.1080/02568543.2011.605208
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). Thousand Oaks, CA: Sage.
- McKay, D. (2013). American politics and society (8th ed.). Somerset, NJ: John Wiley & Sons. Retrieved from <http://www.ebrary.com>
- Metz, A., Halle, T., Baartley, L., & Blasberg, A. (2013). The key components of successful implementation. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.),

Applying implementation science in early childhood programs and systems (pp. 21–42). Baltimore, MD: Paul H. Brookes.

Mims, S. U., Scott-Little, C., Lower, J. K., Cassidy, D. J., & Hestenes, L. L. (2008).

Educational level and stability as it relates to early childhood quality: A survey of early childhood program directors and teachers. *Journal of Research in Childhood Education*, 23, 227–237.

Ministry of Education in People's Republic of China. (2001). *Guidelines for kindergarten education* (trial version). Retrieved from <http://baike.baidu.com/view/2530944.htm>

Mitchell, A. (2009). *Quality Rating and Improvement Systems as the framework for early care and education system reform*. The BUILD Initiative.

Mitchell, A. (2012). *Approaches to financial incentives in QRIS: Approaches and effects*. Retrieved from <http://www.qrisnetwork.org/sites/all/files/resources/gscobb/2012-05-24%2015:13/Approaches%20to%20Financial%20Incentives%20in%20QRIS.pdf>

Mitchell, A., & Mathias, D. (2014). *Re-alignment and re-purposing: How can states maximize existing funding to support cross-sector QRIS?* Webinar PPT. Retrieved from: <http://qrisnetwork.org/lt/2014-financing-quality-through-quality-rating-and-improvement-systems-qris-learning-table/session-6>

Moore, C. A., & Morris, A. (1967). *The Chinese mind: Essentials of Chinese philosophy and culture*. Honolulu, HI: East-West Center Press.

- NAEYC Public Policy Fact Sheet. (2010). *North Carolina state documents regulating health and safety in child care*. Retrieved from <http://nrckids.org/index.cfm/resources/state-licensing-and-regulation-information/north-carolina-regulations/>
- National Association for the Education of Young Children. (2011, September). *From the NAEYC Public Policy Program, as adopted by the NAEYC Governing Board*. Retrieved from https://www.naeyc.org/files/naeyc/2011_QRIS_Statement_0.pdf
- Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, 46, 532–566.
- NICHD Early Child Care Research Network. (2000). Characteristics and Quality of Child Care for Toddlers and Preschoolers. *Applied Developmental Science*, 4(3), 116–135. doi:10.1207/S1532480XADS0403_2
- North Carolina Child Care Resources & Referral Council. (n.d.). *Support for working families*. Retrieved from <http://childcarerrnc.org/s.php?subpage=ChallengesforWorkingFamilies>
- NVivo 10 for Windows. <http://download.qsrinternational.com/Document/NVivo10/NVivo10-Getting-Started-Guide.pdf>
- Office of Planning, Research and Evaluation. (2010). *North Carolina Star Rated License System: QRS Profile*. Washington, DC: Author. Retrieved from http://mathematica-mpr.org/publications/pdfs/earlychildhood/north_carolina.pdf
- Ogbu, J. U. (1988). Cultural diversity and human development. *New Directions for Child and Adolescent Development*, 24, 11–28. doi:10.1002/cd.23219884203

- Padgett, D. (2012). *Qualitative and mixed methods in public health*. Thousand Oaks, CA: Sage.
- Pan, Y., Liu, Y., & Lau, E. Y. H. (2010). Evaluation of the Kindergarten Quality Rating System in Beijing. *Early Education and Development, 21*, 186–204.
- Paulsell, D., Tout, K., & Maxwell, K. (2013). Evaluating implementation of quality rating and improvement systems. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 269–293). Baltimore, MD: Paul H. Brookes.
- Pei, X. (2010). *The regional culture study of Chinese early childhood education under globalization: The cultural appropriateness of early childhood education* (Doctoral dissertation). Retrieved from East China Normal University Dissertation Database.
- Perlman, M., Zellman, G. L., & Le, V. N. (2004). Examining the psychometric properties of the early childhood environment rating scale-revised (ECERS-R). *Early Childhood Research Quarterly, 19*, 398–412.
- Pianta, R. C., Hamre, B., & Stuhlman, M. (2003). Relationships between teachers and children. *Handbook of Psychology*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/0471264385.wei0710/full>
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144–159.

- Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, 102, 225–238.
- Pounder, K. W. (2010). *Early childhood education and North Carolina's Smart Start Initiative*. Retrieved from <https://iei.ncsu.edu/wp-content/uploads/2013/02/Ponder-Early-Childhood-Ed-and-Smart-Start.pdf>
- Quah, S. R., & Quah, J. S. T. (1988). *Singapore*. Santa Barbara, CA: Clio Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=38958>
- Sabol, T. J., & Pianta, R. C. (2014). Validating Virginia's quality rating and improvement system among state-funded pre-kindergarten programs. *Early Childhood Research Quarterly*, 30(Part B), 183 – 198. Retrieved from <http://www.sciencedirect.com/science/article/pii/S088520061400026X>
- Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly*, 18, 427–445. doi:10.1016/j.ecresq.2003.09.004
- Schuyler Center for Analysis and Advocacy. (2012). Quality: What it is and why it matters in early childhood education. Retrieved from: http://www.scaany.org/documents/quality_earlyed_scaapolicybrief_sept2012.pdf

- Schulman, K., Matthews, H., Blank, H., & Ewen, D. (2012). *A count for quality: Child care center directors on rating and improvement systems*. Report from The National Women's Law Center and CLASP. Retrieved from <http://www.nwlc.org/resource/count-quality-child-care-center-directors-rating-and-improvement-systems>
- Sha, L., & Yao, Y. (2014). Lesson learned about teacher-child ratio from major countries of the world: Based on the comparison and analyses of data from 10 countries. *Comparative Education Review*, 5, 35–39.
- Siong, N. B., & Chen, G. (2007). *Dynamic governance: Embedding culture, capabilities and change in Singapore (English Version)*. Singapore, SGP: World Scientific & Imperial College Press. Retrieved from <http://www.ebrary.com>
- Smith, A. B., McMillan, B. W., Kennedy, S., & Ratcliffe, B. (1988). The effect of improving preschool teacher/child ratios: An “experiment in nature.” *Early Child Development and Care*, 41, 123–138. doi:10.1080/0300443880410111
- Solomon, C., & Schell, M. S. (2009). *Managing across cultures: The 7 keys to doing business with a global mindset*. McGraw-Hill, Inc.
- Stoney, L. (2014). *Financing strategies: Helping ECE providers and practitioners secure the funding needed to succeed in QRIS*. Webinar. Retrieved from <http://qrisnetwork.org/lt/2014-financing-quality-through-quality-rating-and-improvement-systems-qris-learning-table/session-5>
- Talan, T. N., & Bloom, P. J. (2004). *Program administration scale: Measuring early childhood leadership*. New York, NY: Teachers College Press.

- Tietze, W., Cryer, D., Bairrão, J., Palacios, J., & Wetzel, G. (1996). Comparisons of observed process quality in early child care and education programs in five countries. *Early Childhood Research Quarterly*, 11, 447–475. doi:10.1016/S0885-2006(96)90017-4
- Tobin, J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in three cultures revisited: China, Japan, and the United States*. Chicago, IL: University of Chicago Press.
- Tobin, J. J., Wu, D. Y. H., & Davidson, D. H. (1989). *Preschool in three cultures: Japan, China, and the United States*. New Haven, CT: Yale University Press.
- Tout, K., Starr, R., Isner, T., Cleverland, J., Soli, M., & Quinn, K. (2010). *Evaluation of Parent Aware: Minnesota's quality rating and improvement system pilot*. Child Trends. Retrieved from https://s3.amazonaws.com/Omnera/VerV/s3finder/38/pdf/Parent_Aware_Year_3_Evaluation_Report_Nov_2010.pdf
- Tout, K., Starr, R., & Soli, R. (2010). *Child care quality rating system (QRS) assessment: Compendium of quality rating systems and evaluation*. Retrieved from http://www.childtrends.org/our-research/early-childhood-development/?view_all=1&topic_area=early-childhood-development#sthash.6qKIM2yi.dpuf
- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations*. Mathematica Policy Research.
- Wang, G., & Zheng, Y. (2013). *China development and governance*. Singapore and Hackensack, NJ: World Scientific Pub. Co. Retrieved from <http://lib.myilibrary.com?id=416340>, accessed November 26, 2014

- Wells, M. B. (2014). Predicting preschool teacher retention and turnover in newly hired Head Start teachers across the first half of the school year. *Early Childhood Research Quarterly*, 30(Part B), 152–159. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0885200614001148>
- Whitebook, M. (2003). *Early education quality: Higher teacher qualifications for better living environments—A review of the literature*. Berkeley, CA: Center for the Study of Child Care Employment.
- Whitebook, M., Kipnis, F., Sakai, L., & Austin, L. J. (2012). Early care and education leadership and management roles: Beyond homes and centers. *Early Childhood Research & Practice*, 14(1).
- Whitebook, M., Sakai, L. M., & Howes, C. (2004). Improving and sustaining center quality: The role of NAEYC accreditation and staff stability. *Early Education and Development*, 15, 305–326.
- Whitehead, R. L., & Whitehead, R. M. (1978). *China, search for community*. New York, NY: Friendship Press.
- Wu, Y. (2013). An overview of Singapore Pre-school Accreditation Framework. *Journal of World Education*, 330, 55–58
- Yang, C. (2004). Lesson learned from the modernization of early childhood education. *Sichuan Normal University Academic Journal*, 6, 126–130.
- Yazejian, N., & Iruka, I. U. (2014). Associations among tiered quality rating and improvement system supports and quality improvement. *Early Childhood*

- Research Quarterly*, 30(Part B), 255–265. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0885200614000581>
- Yin, R. K. (2009). *Case study research: Design and methods*. Los Angeles, CA: Sage.
- Zaslow, M. J., Martinez-Beck, I., Tout, K., & Halle, T. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes.
- Zellman, G. L., & Fiene, R. (2012). *Validation of quality rating and improvement systems for early care and education and school-age care*. Research-to-Policy, Research-to-Practice Brief. OPRE 2012-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from <http://files.eric.ed.gov/fulltext/ED534457.pdf>
- Zhu, J., & Zhang, J. (2008). Contemporary trends and developments in early childhood education in China. *International Journal of Early Years Education*, 28, 173–182.

APPENDIX A

SURVEY

Satisfaction and Effectiveness Survey on the NC Star Rated License System

In this survey, we are interested in learning more about your thoughts, feelings, and ideas on the North Carolina Star Rated License system. When answering these questions, please consider your most recent experience with the NC Star Rated License system. Your answers will be used as a whole group and anonymous, and therefore we will not be evaluating individual responses. As such, please be as honest as possible - there are no right or wrong answers.

Section I. – Knowledge and Satisfaction with the NC Star Rated License System

Questions in this section ask about your overall knowledge and satisfaction with the NC Star Rated License system. All the questions need to be answered on a scale from 1 (low) to 5 (high):

1. How much do you know overall about North Carolina's Star Rated License system?	1	2	3	4	5
	No knowledge	Little knowledge	Some knowledge	A good amount of knowledge	A lot of knowledge

2. How much do you know about the following components of the NC Star Rated License system:

	No knowledge	Little knowledge	Some knowledge	A good amount of knowledge	A lot of knowledge
a. The state requirements for teacher-child ratio and group size for the Star Rated License	1	2	3	4	5

	No knowledge	Little knowledge	Some knowledge	A good amount of knowledge	A lot of knowledge
b. The state requirements for staff qualifications for the Star Rated License	1	2	3	4	5
c. The assessment process for the Star Rated License (e.g., when ITERS-R, ECERS-R, etc. are completed)	1	2	3	4	5
d. The feedback you received after the quality rating	1	2	3	4	5
e. The opportunities available to help increase your stars (i.e., quality improvement through continuous education trainings, workshops, mentoring/coaching, etc.)	1	2	3	4	5
f. The financial incentives (e.g., subsidies, scholarships, wage enhancements, etc.) that NC provides related to the Star Rated License	1	2	3	4	5

3. How satisfied are you with the following components of the NC Star Rated License system:

	Not satisfied	Slightly satisfied	Somewhat satisfied	Satisfied	Very satisfied
a. The state requirements for teacher-child ratio and group size for the Star Rated License	1	2	3	4	5
b. The state requirements for staff qualifications for the Star Rated License	1	2	3	4	5
c. The assessment process for the Star Rated License (e.g., when ITERS-R, ECERS-R, etc. are completed)	1	2	3	4	5
d. The feedback you received after the quality rating	1	2	3	4	5
e. The opportunities available to help increase your stars (i.e., quality improvement through trainings, workshops, mentoring/coaching, etc.)	1	2	3	4	5
f. The financial incentives (e.g., subsidies, scholarships, wage enhancements, etc.) that NC provides related to the Star Rated License	1	2	3	4	5

4. How much do you think **parents/guardians** know about the following components of the NC Star Rated License system

	No knowledge	Little knowledge	Some knowledge	A good amount of knowledge	A lot of knowledge
a. The star level of your program	1	2	3	4	5
b. The state requirements for teacher-child ratio and group size for the Star Rated License?	1	2	3	4	5
c. The state requirements for staff qualifications for the Star Rated License?	1	2	3	4	5
d. The assessment process for the Star	1	2	3	4	5

	No knowledge	Little knowledge	Some knowledge	A good amount of knowledge	A lot of knowledge
Rated License (e.g., when ITERS-R, ECERS-R, are completed)					
e. The opportunities available to help increase your stars (i.e., quality improvement through trainings, workshops, mentoring/coaching, etc.)	1	2	3	4	5
f. The financial incentives (e.g., subsidies, scholarships, wage enhancements, etc.) that NC provides related to the Star Rated License	1	2	3	4	5

5. How much does the
Star Rated License
affect your work?
- 1 No
impact at
all
- 2 Slight
impact
- 3 Has some
impact
- 4 A good
amount
of impact
- 5 A lot of
impact
6. How effective is the
Star Rated License in
enhancing program
quality in your
program?
- 1 Not
effective
at all
- 2 Slightly
effective
- 3 Somewhat
effective
- 4 Effective
- 5 Very
effective
7. How much stress does
the Star Rated License
process create for you?
- 1 No stress
at all
- 2 Slight
stress
- 3 Some
stress
- 4 A good
amount
of stress
- 5 A lot of
stress

Section II. Perceived Accuracy of Quality Ratings

Questions in this section ask specifically about the process of assessment.

8. How many times have you experienced the Star Rated License process since you started working **in this program**? _____
9. How many times have you experienced the Star Rated License process since **you started your career** in early childhood? _____
10. Rate the accuracy that you perceived for the result of your quality rating:

	Not accurate at all	Slightly accurate	Somewhat accurate	Pretty accurate	Very accurate	Not applicable
a. To what degree do you think your current Star Rating for your program is accurate?	1	2	3	4	5	N/A
b. To what degree do you think your last ECERS-R assessment was accurate?	1	2	3	4	5	N/A

Section III. Perceived Effectiveness of Quality Improvement

Questions in this section ask specifically about the quality improvement you received before your last Star Rated License assessment process began.

11. Overall, how much do you think the quality improvement activities you received (e.g., professional development training, workshops, coaching, etc.) were effective in improving your classroom/program quality?
- | | | | | | |
|----------------------|--------------------|--------------------|-----------|----------------|----------------|
| 1 | 2 | 3 | 4 | 5 | N/A |
| Not effective at all | Slightly effective | Somewhat effective | Effective | Very effective | Not applicable |

12. What kinds of support did you receive before the Star Rated License assessment process began? Select all that apply.

- a. continuous education
- b. one-to-one coaching/mentoring
- c. workshops or meetings
- d. resource sharing with peers, colleagues or other programs
- e. other _____ (specify what it is/they are)

13. How effective were the quality improvement activities that are listed below:

	Not effective at all	Slightly effective	Somewhat effective	Effective	Very effective	Not applicable
a. Continuous education	1	2	3	4	5	N/A
b. One-to-one coaching/mentoring	1	2	3	4	5	N/A
c. Workshops or meetings	1	2	3	4	5	N/A
d. Resource sharing with peers, colleagues or other programs	1	2	3	4	5	N/A
d. other (specify):	1	2	3	4	5	N/A

Demographics

1. Age_____
2. Gender: ☐Female ☐Male
3. Race and ethnicity (check all that apply):
☐ African American
☐ Asian/Pacific islander
☐ White/European American
☐ Native American
☐ Other
☐ Hispanic; ☐ Non-Hispanic
4. Highest education level:
☐ less than 2-year degree
☐ 2-year college degree;
☐ 4-year degree in Early Childhood Education/ Child Development;
☐ 4-year degree in other major;
☐ Graduate degree
5. The year that the highest degree was earned_____
6. Years of experience in Early Childhood Education (Paid employment only):

7. Length of time employed in this program: _____years _____months
8. Which best describes your program (check all that apply):
☐For-profit; ☐Not-for-profit; ☐Public-sponsored;
☐Cooperate-sponsored; ☐Religious-affiliation; ☐Head Start;
 Other: _____
9. Current program rating ☐5-star; ☐4-star; ☐3-star; ☐2-star; ☐1-star
10. Your position in the program:
☐ Director
☐ Lead Teacher
☐ Teacher assistant

Thank you for completing this survey.

APPENDIX B

ADMINISTRATOR INTERVIEW

Section I. Background question

1. Could you describe QRIS in your state/district/country
 - 1) History: when it launched, how it has evolved over time
 - 2) What policies/standards are used to align QRIS?
 - National
 - Local
 - Other
 - 3) What is the time frame for the QRIS? How long does each aspect take to complete?

Section II. Components in QRIS

2. What are the components included in the R (Rating)?
 - 1) What is included in the QRIS and/or licensing process?
 - a. What structural components (e.g., ratio, group size, teacher/administrator education, experience, financial status, health/safety, fire, playground safety etc.)
 - b. What measurement tool or set of tools are used in your state/district/country to **rate quality**?
 - 2) Who conducts the ratings/assessments and who gathers the structural components (are they separate)?
 - a. Who are the staff that are assessing ECE program?
 - b. What is the criteria/process for selecting assessment staff members to implement QRIS?
 - c. What is the working status of the assessors?
 - i. Are they working part time/full time?
 - ii. Is there high turnover in these positions or does staff remain in their roles for long periods of time? Why?
 - d. Is there training program available for the assessors?
 - i. How do the assessors get reliable?
 - ii. How are the assessors evaluated?

- iii. Are there opportunities for training or education for the assessors? What types?
- 3. What are included in the I (improvement)?
 - 1) When does it take place?
 - before, during, after the assessment?
 - 2) What are the general approaches of improvement?
 - technical assistance, coach, workshops, mentoring, etc.
 - 3) Who are the technical assistants?
 - a. Who are the staff that are supporting program improvement?
 - Coaching
 - Consulting
 - Coordinating – arranging all these activities
 - b. What is the criteria/process for selecting staff members to implement QRIS?
 - Coaching
 - Consulting
 - Coordinating
 - c. What is the working status of the staff?
 - i. Are they working part time/full time?
 - ii. Is there high turnover in these positions or does staff remain in their roles for long periods of time? Why?
 - d. Is there training program available for the staff?
 - i. How are the staff evaluated?
 - ii. Are there opportunities for training or education for the staff? What types?
- 4) How do you assess fidelity (effectiveness of the improvement; for example, if the program's quality get improved in the next assessment after implementing the improvement activities)?
 - a. How do you know the implementation is effective?
 - i. Do you have established data resources?
 - b. How are the assessment results utilized?
 - i. Do you use assessment result to help advise programs
 - ii. Do you analyze assessment result to inform change?

Section III. Values

4. Is the QRIS valued in this city/country?
 - 1) In what ways?
 - 2) How is it valued?
5. What do you perceive as the value for programs, teachers, and families?
 - 1) How do you promote buy-in?
 - a. In programs
 - b. In teachers
 - c. In families
6. From your perspective what is the most valuable component in the QRIS?
7. What problems are relatively difficult but important to solve in your opinion?
(1.2.6.3)
 - 1) What problems are usually seen in the implementation of QRIS?
 - 2) What are the general approaches to address the problems?
 - 3) From your perspective what is the component that most needs enhancement in the near future?
8. From your perspective, how do the R and I go hand-in-hand?
 - How independent and interdependent are the two components?

Is there anyone that you want to refer to provide additional information or perspectives?

APPENDIX C

CULTURAL CROSS-COMPARISONS

QRIS Components Cross-comparison

	Standards	Measurement	Improvement	Financial Incentives	Consumer Education
North Carolina	Two components (7 points): teacher education and program standards (ratio & group size, center environment) + bonus (1 point): family partnership, or administration and management	ERS (ECERS-R, ITERS-R, FCCERS & SACERS) Outsider assessment: observation and interview	Consultants from DCDEE, TA providers from Smart Start, Resource & Referral (R&R)	Subsidy for low-income families and children with special needs. Only 3-, 4-, and 5-star programs can receive subsidy.	Parents can search for childcare programs on DCDEE's website.
Singapore	Seven components: leadership, planning and administration, staff management, resources, curriculum, pedagogy, and health, hygiene and safety	Quality Rating Scale (QRS) Self-appraisal and outsider assessment: observation and interview	Facilitation from assessors, consultants, and district directors	Not applicable	Parents are informed about the new system. Parents may be required to cooperate on assessment day.
Beijing	Basically including structural quality (physical environment, such as size of building, equipment and materials, and staff qualification), and process quality (Administration, Education and Pedagogy, Health and Hygiene, and Child Assessment). Different	Non-standardized measurement. Three-level assessment: self-appraisal + experts evaluation (for Demonstration program only) + Community evaluation	On-site instruction and facilitation; Assigned professional development opportunities	Not applicable	Parents know that "Demonstration" means the best, and 1R1C is also very good.

Standards	Measurement	Improvement	Financial Incentives	Consumer Education
standards for different quality levels.	(for non-demonstration programs only)			

QRIS Implementation Cross-comparison

	Competency Driver		Organization Driver			Leadership	Stage
	Staff selection	Staff training & coaching	Facilitative administration	System intervention	Data-driven decision-making		
North Carolina	Assessors: education level (bachelor); working experience in ECE field; Based on candidates' reliability scores in using ERS. Consultants and TA providers: education level (bachelor); working experience in ECE field	Training are integrated in selection and work; Reliability is supervised and checked regularly	Leaders are commit to their leading positions; Communication and cooperation between assessor and TA provider is weak; Great ability in making adjustments in policies and practices	Cooperation with research institutions; Cooperation with agencies (Smart Start, R&R); cooperation between the advisory committee and DCDEE	Used for staff selection and training; Look for evidence of the effectiveness of the system, including measurement validity and fidelity of the whole rating system in differentiate quality levels	Facilitative leader; Form coalition and partner relationships with leaders in other domains.	Full implementation; In need of new adjustment in several places (goes back to earlier stages)
Singapore	Retired elementary school principals; Based on reliability scores in using QRS	First time: 100 credit hours ECE course; 2-day training on QRS in childcare center and kindergarten respectively; Annual standardized training	One primary leader and several within system leaders; Promote buy-in in programs; Great capability in adjusting policies and practices.	Little cooperation with external institutions or agencies	Three pilot studies with programs before launch; Reliability check in the first round of staff training; Reexamine measurement validity after five years'	Directory and participative leadership combined	Initial implementation stage; Moving fast from exploration to the current stage

	Competency Driver		Organization Driver			Leadership	Stage
	Staff selection	Staff training & coaching	Facilitative administration	System intervention	Data-driven decision-making		
Beijing	Assessors: Retired experts of different specialties in ECE; Professional development coaches: master teachers from different kindergartens	Professional development coaches engage in research to maintain their professionalism	Committed leader; Learning Community promotes frequent communication among programs, and harmonious organizational climate	Cooperation with professional development institution.	No formal data base	Directory leadership with some decentralized responsibilities to senior staff (i.e., demonstration program directors)	Hard to define using criterion of implementation science; Roughly at the full implementation stage in terms of its stability, but not its components in each driver.