

**A PHENOMENOGRAPHIC STUDY OF SINGAPORE CHINESE
LANGUAGE TEACHERS' EXPERIENCE OF ASSESSMENT TENSION**

CHAN KWONG TUNG

NATIONAL INSTITUTE OF EDUCATION
NANYANG TECHNOLOGICAL UNIVERSITY

2019

**A Phenomenographic Study of Singapore Chinese Language
Teachers' Experience of Assessment Tension**

Chan Kwong Tung

A dissertation submitted to the
National Institute of Education,
Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor in Education

2019

Acknowledgments

To the ones I love, care and respect.

To my dearest mother, Wong Wan Heung, who left us much too soon.

I am sure you can see and feel me, Mother - You will be so proud of me.

This research would not have been possible without the help and support of many people. I have had the honor and privilege to learn from my supervisor and supportive faculty, understanding colleagues, and the best group of classmates a part-time student could ever ask for.

Mere words are simply not enough to express my gratitude and appreciation that I have for my supervisor, A/P Tan Heng Kiat, Kelvin. A/P Tan's insight, patience, guidance and coaching were invaluable. Special thanks to A/P Kim Koh of the Werklund School of Education, University of Calgary for her supervision during the early stage of my EdD study. In addition, I also extend my gratitude to Dr Karen Lam, Dr Rachel Goh and Ms Lin Rongchan for their insightful comments during the last stage of my writing.

I am most thankful for the assistance given by the faculty and staff from the Curriculum, Teaching and Learning (CTL) Academic Group of the National Institute of Education (NIE), Nanyang Technological University (NTU). They have provided me with valuable advice and support at different stages of my study and research. I would also like to thank the three schools and the nine teachers who have participated in this research. Without their participation, this research would not have been completed. I appreciate their time and effort during the interview period.

My family has always been an important source of love and inspiration. I am indebted to my family for their unconditional love and patience in this pursuit: my dearest wife, Yvonne, who had sacrificed so much throughout my research, when I had to juggle work with study; my wonderful three daughters, Genevieve, Aurora and Minerva for their understanding and sacrifice; and my father (Cheng Chuen) and parents-in-law (Johnny and Rose), who have been taking care of the family. Without the stability and anchor provided by the love and care of my family, this research would not have been possible.

Table of contents

Acknowledgments.....	i
Table of contents.....	ii
Abstract.....	v
List of Figures.....	vii
List of Tables.....	viii
List of Abbreviations.....	ix
CHAPTER 1 : INTRODUCTION	1
1.1 Assessment complexity in reality: Tension experienced by teachers.....	1
1.2 Assessment in Singapore.....	2
1.3 Situating Chinese Language (CL) teaching and assessment in the Singapore context	10
1.4 The nature of assessment tension (AT) and the research question.....	14
1.5 Significance of the research.....	18
1.6 Research method.....	21
1.7 Summary.....	21
CHAPTER 2 : REVIEW OF LITERATURE	24
2.1 Introduction.....	24
2.2 Literature search method.....	24
2.3 AT comprising different assessment purposes.....	26
2.4 AT and “two contrasting paradigms of assessment”	30
2.5 AT caused by new assessment reforms and policies.....	34
2.6 AT as negative and undesirable	37
2.7 Singapore teachers’ experience of AT.....	40
2.8 AT and the history and culture of assessment.....	42
2.9 Summary.....	48
CHAPTER 3: RESEARCH METHODOLOGY	51
3.1 Introduction.....	51
3.2 Research methods commonly used in the studies of AT.....	53
3.2.1 The issues with quantitative research approach.....	54

3.2.2 The issues with case study.....	56
3.3 Unique contributions of phenomenography to explore teachers’ experience of AT.....	59
3.3.1 Basic concepts of phenomenography.....	61
3.3.2 Unique features of phenomenography.....	62
3.3.3 Aims of research and sample variation.....	63
3.3.4 The structure of awareness and outcome space.....	65
3.4 Data sources and collection.....	68
3.4.1 Identifying participants.....	68
3.4.2 Planning and conducting the interviews.....	70
3.5 Data analysis.....	78
3.6 Role of the researcher.....	81
3.7 Validity, reliability and “bracketing”.....	83
3.8 Summary.....	87
CHAPTER 4: FINDINGS	88
4.1 Introduction.....	88
4.2 The outcome space.....	90
Conception A (Passive).....	93
Conception B (Responsive).....	94
Conception C (Propulsive).....	95
4.3 Qualitative difference in teachers’ experience of AT.....	98
4.3.1 Conception of AT in terms of purposes.....	100
4.3.2 Conception of AT in terms of reality and ideal gap.....	105
4.3.3 Conception of AT in terms of involvement of stakeholders of education.....	112
4.4 Summary.....	117
CHAPTER 5: DISCUSSION AND CONCLUSION	119
5.1 Recapitulation of research purpose and findings.....	119
5.2 Significance and contributions of the research.....	119
5.2.1 Understanding AT to be positive and constructive.....	120
5.2.2 Increasingly importance of FA in a SA-dominated educational landscape.....	123
5.2.3 Re-looking AT in sustainable assessment.....	128
5.2.4 Re-discovering the open tradition of Confucian education from CL teachers’ experience of AT.....	134

5.3 Implications and recommendations for practice and policy.....	139
5.4 Limitations of the research and suggestions for further research.....	147
5.5 My experience of phenomenographic research.....	148
5.6 Summary.....	150
REFERENCES	152
APPENDICES	187
Appendix A: Glossary of Assessment Concepts (Purpose of Assessment).....	187
Appendix B: Literature Reviews on Assessment Tension.....	188
Appendix C: Sample invitation email to participants.....	193
Appendix D: Attached email documents.....	194
Appendix E: Interview Protocol.....	198
Appendix F: Sample of preliminary categories of description.....	200
Appendix G: Proposed conceptions of assessment tension.....	203
Appendix H: Original Extracts in Chinese.....	204

Abstract

This research investigates the qualitatively different ways in which Chinese Language (CL) teachers in Singapore experience assessment tension (AT). AT is defined as “teachers’ feelings of struggle with competing considerations in their assessment practices”. A phenomenographic approach was used to identify and describe CL teachers’ collective experience of AT in terms of their distinctive differences and their interrelatedness within. Data was collected from the semi-structured interviews.

In the literature, teachers are increasingly experiencing much more AT at different levels for the past two decades due to a variety of reasons, both in Singapore and elsewhere. This research investigates specifically what assessment considerations CL teachers focus on and how they deal with these competing considerations. The investigation focuses on the different ways CL teachers described their AT experience. The primary research question for this research is: “What are the qualitatively different ways in which teachers experience AT?”

A total of nine CL teachers from three secondary schools of different profiles and the Ministry of Education (MOE) headquarters were selected as participants. The findings or the outcome space of the investigation could be logically reduced into the ‘What’ aspect and the ‘How’ aspect of the phenomenon experienced. It may be understood in three related and distinct ways.

At the collective level, the qualitatively different ways in which teachers experience AT are denoted as “conceptions” and are represented by “categories of description”. The three conceptions of experiencing AT found in this research are the passive, the responsive and the propulsive. Respectively, their focal awareness of their experience of the phenomenon was on current limitations, imminent opportunities and beyond current limitations and imminent opportunities (or beyond school). Correspondingly, teachers assume the role of either a compliant executor, a rational practitioner or an advanced re-conceptualiser in their assessment practices. The typical attributes of these conceptions are described and their relationships are uncovered in this research. The outcome space of CL teachers’ progressive awareness of AT is further analysed in terms of the assessment purpose, reality and ideal gap, demand from

stakeholders of education. These three dimensions of variation, as well as the attributes of the three conceptions of AT, are described in detail.

The variation of teachers' experience of AT and the structure of awareness of AT derived from teachers' interviews are useful for re-addressing assessment issues that include the understanding AT beyond a negative connotation, and the increasing importance of formative assessment (FA) in a summative assessment (SA)-dominated educational landscape. The findings also allow readers to re-look at AT in sustainable assessment practice and re-assess the taking up of new assessment ideas from the 'West' in a Confucian-heritage context/culture (CHC). More importantly, the progressive awareness of AT sheds light on the theoretical, analytical and pedagogical aspects of this phenomenographic study. The structure of awareness of AT depicts how the experience of assessment and education policies is theorised by individuals (theoretical aspect), is analysed to present the different aspects of ways of experiencing AT and how these aspects are related (analytical aspect). The distinctive research benefit of knowing teachers' ways of experiencing AT is that policy-makers and school leaders could suggest how 'better' ways of understanding and experiencing AT may be brought about (pedagogical aspect). Consequently, this outcome space would also reveal further implications on how AT may be understood in terms of assessment practice and policy making.

Key words: assessment tension, Chinese Language, phenomenography, conceptions of assessment

List of Figures

Figure 3.3.4: Structure of awareness

Figure 3.5: Process of phenomenographic data analysis

Figure 5.2.2a: Focal awareness and thematic field of teachers in Conception A

Figure 5.2.2b: Focal awareness and thematic field of teachers in Conception B

Figure 5.2.3: Focal awareness and thematic field of teachers in Conception C

Figure 5.3: The structure of awareness of AT

List of Tables

Table 1.3: AT at the policy and practice levels

Table 2.4: Two Contrasting Paradigms of Assessment

Table 3.3.4: Structure of awareness

Table 3.4.2: Types of interview questions used in a single interview

Table 4.2: Teachers' qualitatively different ways of experiencing AT

Table 4.3: The three conceptions of AT

Table 5.2.2a: Past relationships between FA and SA

Table 5.2.3a: Comparison between sustainable assessment and Conception C

Table 5.2.3b: Sustainable assessment practices in tension

Table 5.3a: The three aspects of phenomenographic study

Table 5.3b: Agendas between short- and long-term learning

List of Abbreviations

- Academy of Singapore Teachers (AST), 10
- assessment for learning (AfL), 1
- assessment of learning (AoL), 1
- assessment tension (AT), 1
- Chinese Language (CL), 1
- Confucian-heritage contexts/cultures (CHCs), 3
- Continual Assessment (CA), 2
- Cultural–historical Activity Theory (CHAT), 40
- curriculum and standards framework (CSF), 36
- Curriculum Planning and Development Division (CPDD), 9
- Curriculum, Teaching and Learning (CTL), 85
- Design and Technology (D&T), 35
- Doctor in Education (EdD), 85
- Education Resources Information Center (ERIC), 24
- English as Second Language (ESL), 38
- English Language (EL), 40
- Enhanced Performance Management System (EPMS), 4
- Exploratory Factor Analyses (EFA), 55
- formative assessment (FA), 1
- formative use of summative testing (FUST), 45
- General Certificate of Education (GCE), 2
- General Certificate of Secondary Education (GCSE), 33

Head-of-Department (HOD), 9

Holistic Assessment (HA), 5

Information and Communication Technology (ICT), 11

Institutional Review Board (IRB), 69

Key Performance Indicators (KPIs), 4

Knowledge Building (KB), 40

Massachusetts Comprehensive Assessment System (MCAS), 38

Ministry of Education (MOE), 1

Mother Tongue Language (MTL), 2

Mother Tongue Languages Review Committee (MTLRC), 12

Nanyang Technological University (NTU), 64

National Assessment Program for Literacy and Numeracy (NAPLAN), 36

National Institute of Education (NIE), 10

National University of Singapore (NUS), 64

Primary Education Review and Implementation (PERI), 5

Primary English Assessment for Learning (PEAL), 43

Primary School Leaving Examinations (PSLE), 2

School-based assessment (SBA), 44

Singapore Centre for Chinese Language (SCCL), 10

Singapore Examinations and Assessment Board (SEAB), 10

summative assessment (SA), 1

Survey of Assessment Beliefs (SAB), 54

Teach Less Learn More (TLLM), 6

Teachers' Conceptions of Assessment (TCoA), 53

Teachers' Classroom Assessment Scale (TCAS), 55

Thinking Schools, Learning Nation (TSLN), 6

CHAPTER 1

INTRODUCTION

This chapter is an introduction to my research into Chinese Language (CL) teachers' experience of assessment tension (AT). The core educational problem, the nature of AT and the overarching research questions are presented. The context and the significance of the research are also described, with a brief mention of the research method. This chapter concludes with an overview of the thesis.

1.1 Assessment complexity in reality: Tension experienced by teachers

Although there are technical definitions for every assessment term and practice, educational assessment is seldom straightforward. Assessment is commonly depicted in the ways defined by scholars and researchers. Assessment terms such as formative assessment (FA), summative assessment (SA), assessment for learning (AfL), assessment of learning (AoL) are very neatly categorised. In my context, the Ministry of Education (MOE) has laid out these definitions of assessment in the intranet portal site. The detailed definitions are shown in Appendix A. For example:

Formative Assessment is carried out during the instructional process to provide feedback to adjust ongoing teaching and learning in order to improve the students' achievement of intended instructional outcomes.....Summative Assessment is carried out at the end of an instructional unit or course of study for the purpose of giving information on students' mastery of content, knowledge and skills, assigning grades or certifying student proficiency. (MOE-OPAL, 2018)

These terms (such as FA, SA, AfL, AoL) seem non-contentious and serve their respective specific purposes. However, assessment in practice may not be so clear-cut and may also need to do multiple-duties, which gives rise to a lot of tension because teachers are worried that their assessments are not able to fulfil all these duties. These assessment experiences may turn out to be highly challenging and emotional for teachers. Although MOE has laid out clear definitions for FA and AfL, they are often used loosely by teachers

on the ground. For this reason, FA and AfL has also been used interchangeably in this dissertation.

Leong, Cheng and Tan (2015) recognise that depending on the primary sub-fields of education, assessment can have some differences in “referring names” and meanings (p.1). These names demonstrate that assessment in practice may not be as tidy as the terms suggest. The issue gets more complicated when teachers have different perceptions about assessment. In Bonner’s (2016) words, the story of teacher perceptions about assessment is “a story of tension and conflict” (p.21). Indeed, assessment is far from straightforward because human beings interpret and act on it in an inconsistent, unpredictable and often irrational fashion (Harris & Brown, 2016).

The above tidy technical definitions of assessment, while trying to be succinct and brief, appear to be detached and unaware of the complexity and contradictory ways in which assessments are experienced in reality. These conceptual semantics will remain inadequate if they do not analyse the pragmatic and actual contexts of those who are experiencing the assessment concepts. As Carless (2009) puts it, “assessment is often beset with tensions and compromises” (p. 79). In recognising the complexities and the messy spaces of assessment conceptualisation and actual practices, I believe more records and accounts of how Singapore teachers experience assessment will be valuable. These accounts of experiences span a wide range of practices. For example, how do teachers experience tension in assessment when they set an end-of-year examination paper that is supposed to both evaluate students’ learning, as well as to give feedback for further improvement? What are some of their considerations when they are told to be less “exam-oriented” and shift their practices toward “AfL”? To this end, this research explores assessment from the perspective of the “experiencers”. This will help further unpack the actual meaning of assessment for all stakeholders of education to understand its essence and its implication for education.

1.2 Assessment in Singapore

In Singapore, SA is very prominent in schools. Students' assessment results are recorded into Continual Assessment (CA) and Semestral Assessment reports. Both CA and Semestral Assessment happens throughout the academic year. CA usually consists of class tests, journal entries, quizzes and other exercises formally given by teachers in

class. On the other hand, Semestral Assessment consists of common tests, mid-year and end-of-year examinations. These assessment scores would be recorded and factored into the final computation of students' academic performance. Such scores are determinants of their promotion to the next level of study. Very often, these scores would also decide if a student can be placed into an elite class and be offered the most prestigious and popular subject combinations (such as higher-level Mother Tongue Language or MTL programme, Pure Sciences or Additional Mathematics) in the next academic year.

The national high-stakes examinations such as Primary School Leaving Examinations (PSLE) and Singapore-Cambridge General Certificate of Education (GCE 'O' and 'A' Level examinations) are summative in nature. The PSLE is a yearly national examination which all students sit at the end of their sixth year of primary school education before they move on to secondary schools. The PSLE assesses students' achievement in English, MTL, Science and Mathematics. Students are placed into different streams and schools according to their performance. GCE level examinations are usually taken by students who have completed their four or five years of secondary education and two or three years of their post-secondary education respectively. Again, students would be placed into the next stage of education according to their results.

Any assessments other than CA, Semestral Assessment and national examinations are usually considered formative in nature as marks are not recorded for selection, placement or promotion purposes in schools. I observe that while SA is concrete and clear to teachers, FA is more vaguely defined. Teachers may experience tension between these different competing considerations in terms of assessment purposes and agendas, especially when FA is poorly conceptualised and a common understanding is not held by different stakeholders of education.

The social context in Singapore emphasises meritocracy and highly values educational credentials (Gopinathan, 1997, 2007). This has resulted in a nation-wide obsession with doing well in examinations. Many believe that good examination results will see the child getting into the top schools in the country and eventually help him/her build a brighter future and career prospects. This belief may have its cultural and historical roots. As the majority of Singaporean are of Chinese descent who inherited some Confucian heritage, a few cultural scholars (such as Kuo, 1996; Tan, 2012) have

observed that the Singapore state's vision of shared citizenship (also known as the 'Our Shared Values'), resembles very much the Confucian thinking. According to Tan (2012), Confucianism generally emphasises the importance of the society, community, family, consensus, and harmony. These values form the core of Singapore's 'Shared Values'. It may be these reasons that Singapore has been commonly mentioned by many scholars as a contemporary Confucian-heritage context/culture (CHC), although its population and culture are not entirely Chinese. For example, in education, Carless and Lam (2014) believe that the "exam-oriented" thinking strongly influences how assessment is approached in contemporary CHCs that include Hong Kong and Singapore. They observe that SA dominates to such an extent that it is difficult for FA to be established.

CHCs are commonly known for their "exam- and achievement-oriented" teaching (Luk-Fong & Brennan, 2010), use of repetitive learning (Dahlin & Watkins, 2000) and a high respect for higher authority (Turner, 2011). To this end, CHCs usually favour teacher-centred pedagogy, with an emphasis on academic performance in high-stake examinations, with teachers taking a big responsibility for students' academic results (Luk-Fong & Brennan, 2010). Inevitably, there is much tension between SA and FA in these settings, especially when there is a call for change in assessment practice toward AfL.

It is because of this cultural and historical factor, parents, students, and teachers alike are all drawn into this competitive "exam-oriented" mode (Cheah, 1998). Tan (2010) has further explained the relationship between assessment system and meritocracy in Singapore. He explained that: "The principle of meritocracy directs all stakeholders' attention and effort into the centrally-planned and common assessment framework" (Tan, 2010, p.13). Consequently, it puts a huge pressure on schools that inadvertently demands their teachers to prepare the students well for the national examinations. As a result, teachers become accountable for their students' performance (Cheah, 1998). This is further reinforced by the teachers' appraisal system or the Enhanced Performance Management System (EPMS). The EPMS pits teachers against one another in the same substantive grades. Their performance is assessed relatively to the performance of other teachers within the same grade. In other words, it is a norm-referenced appraisal system. They will compete to be eligible for a performance bonus or promotion opportunities based on a group of Key Performance Indicators (KPIs) and

the ranking is highly competitive. Commonly, teachers in this appraisal system are assessed by the school leaders and key personnel on their ability to ensure success in their students' academic performance, especially in the high-stakes national examinations. Similar to students striving for good examination results, teachers have to compete with their peers in terms of delivering excellent academic results, to be well-appraised in their profession. Hence, assessment in Singapore is also high-stakes for teachers.

As Sharpe and Gopinathan (2002) have commented, achieving results is the bottom line of daily thinking and practice for teachers, giving rise to a number of tensions. Teachers are pressured to prioritise in terms of results and there are further demands from parents, students and even society, which all add complexity to how teachers consider and experience assessment. I, too, observe that teachers often find it hard to reconcile the tension between different purposes and agendas of assessment, especially when there is a call for a greater focus on AfL instead of preparing students well for the AoL.

One good example is the “Holistic Assessment” (HA) initiative recommended by the Primary Education Review and Implementation (PERI) Committee of the MOE in 2009. The two concrete areas the initiative intends to address are the issue of “exam-oriented culture” that includes advocating more qualitative feedback for improvement and the use of “bite-sized” assessment as an alternative to one-off examination(s) (Ratnam-Lim & Tan, 2015, p. 63). It was further found that the contextual factors (such as the “exam-oriented” culture and system of accountability, teacher training, time and supports) may hinder teachers from changing their assessment practice toward FA effectively (Ratnam-Lim & Tan, 2015). These findings reflect my experiences and observations on teachers' assessment practice. Teachers may be caught in situations where they struggle with deciding on the kind of assessment to carry out in their classrooms. For example, some teachers' “exam-oriented” practice may be challenged when there is a call to “balance” or move towards AfL. Such movement inevitably causes them to experience the tension between how to effectively cover and teach the subject syllabus while preparing their students well for the national examinations.

The Assessment Philosophy of MOE has the vision of “Towards Learner-Centred and Balanced Assessment” where assessment should be closely aligned with curricular

objectives, content and pedagogy (MOE, 2012). To a large extent, such an initiative has challenged some teachers' old assessment practices which pay more attention to AoL. The MOE assessment philosophy put forth that:

Both school-based assessment and national examinations play important and different roles in our education system. A balanced assessment system should have both Assessment of Learning as well as Assessment for Learning. (MOE, 2012)

The need for teachers to change their current practices in any new educational policy always brings along a certain degree of tension. This is especially true for Singapore where the education system has been maintaining high standards of student academic achievement in international comparative assessments. Moving into the new century, the challenge is for the education system to think about what it should do to maintain these high standards and simultaneously prepare and support its students well for lifelong learning (Leong & Tan, 2014). To this end, Singapore serves as an ideal context to investigate how teachers face the challenge with the introduction of new classroom assessment initiatives (such as the "Holistic Assessment" and "Balanced Assessment"). Their distinctive experiences of AT deserve stakeholders' attention if they are serious about helping teachers to better manage this tension. The rationale for such assessment initiatives is to address the overemphasis on testing and examination (MOE, 2009). Scholars such as Leong and Tan (2014) have recognised the challenges brought about by the new assessment initiatives in conceptualising FA and implementing changes. I observed that teachers found it hard to reconcile the dilemma of helping students perform in tests and high-stake examinations, and the introduction of formative classroom assessment practice. There are tensions in and between policy and practice when carrying out FA. Leong and Tan (2014) also conclude that the institutional authority of successful high-stake examination results is being challenged under the new policy. Teachers may be too focused on implementing new assessment practices while ensuring their students' high standards performance in high-stake national examinations. However, it is not known how teachers are experiencing this tension and challenge.

In fact, Singapore has been introducing educational reforms that require teachers to change their assessment practices in the recent years and has met with a lot of challenges and issues along the way. One such reform is the “Teach Less Learn More” (TLLM) movement in 2004. TLLM is a policy to realize the “Thinking Schools, Learning Nation” (TSLN) vision. Initiated in 1997, the TSLN aims to transform the education system, and to prepare the students well for the 21st century. TLLM is aimed at reducing syllabus. It would translate to less pressure on students, a bit less rote learning, more space for them to explore and discover their talents and also more space for teachers to think, to reflect, to find ways to bring out the best in their students and to deliver quality results. It is to teach less to our students so that they will learn more (Lee, 2004). Similar to the notion or principle of “Balanced Assessment”, I think it would be too idealistic and simplistic to expect the TLLM to be effected and implemented without much disruption as the complexities of changing teachers’ teaching and assessment practices cannot be ignored. Such policy changes have been assumed to be unproblematic to teachers and students.

Conversely, in Hogan's (2011) account of the TLLM, he gave his views on the relationship between teachers' beliefs and their practices. He concluded that the teachers would probably ignore the new innovations or practices after a while unless they have changed their views in line with the new instructional practices. They were likely to carry on with their old teaching practices in spite of their sincere belief and commitment to the new pedagogical initiatives or reforms. Later, Hogan et al., (2013) reported tension, inconsistency or even misinterpretation and misalignment of what the educational policy and initiative such as the TSLN and TLLM have advocated originally. This empirical finding was drawn on data collected in 2010 to analyse the instructional strategies Singapore secondary three teachers employed in their mathematics and English lessons. Although the subject is different, it makes me remember the experience I had as a CL classroom teacher who had to juggle with educational initiatives or reforms on one hand, and complete the syllabus, bearing a teacher's accountability for students' examination performance on the other. No doubt that tension is present when it comes to assessment decision-making and practices. To this end, such empirical study is valuable because they departed from the idealistic and simplistic views of the initiative offered by the authorities and instead presented the real experiences of teachers.

Tan (2008b) recognised the tension between the qualitative and quantitative aspects of assessment in Singapore. Later, while Tan (2011b) argued that assessment in TLLM initiative should be conducted “more qualitatively, through a wider variety of authentic means, over a period of time to help in their own learning and growth, and less quantitatively through one-off and summative examinations” (as shown on the MOE website, 2007), this was not evident in schools. Tan (2011b) pointed out that the intended “qualitative” approach to assessment was deviated by the pressure from quality assurance such as excellence models and external validations. He further opined that these twisted and shattered students’ learning. Should teachers conform to the initiatives for assessment set by MOE or submit to the current pressure placed on prioritizing students’ results? For such a high-stakes assessment system with its culture of performativity that is also being labelled as a unique CHC where the ‘East meets West’ in a highly globalised country, how do the teachers experience tension in assessment? How do they respond to these challenges?

These are not straightforward questions because individuals would conceptualise, interpret and negotiate assessment according to their own realities. Unsurprisingly, there are so many cases of tension reported in Singapore and elsewhere arising from educational reforms or policy changes (Arkoudis & O’Loughlin, 2004; Dreher, 2012; Ecclestone & Swan, 1999; Hardy, 2013; Kim et al., 2013; Hopfenbeck, Flórez Petour & Tolo, 2015; Kostogriz & Doecke, 2011; Pinto, 2016; Volkmann et al., 2005). These studies concluded that new educational policies or reforms would require teachers to respond and change their teaching and assessment practices in ways which could be very personal and diverse. Regardless of teachers’ choice of responses to the new educational policies or reforms, they experienced different degree of conflict and tension (Cohen & Ball, 1990). Vandeyar (2005), quoting earlier works done by Black (1993), Crooks (1998); Blok, Otter & Roeleveld (2002) and others, reported that teachers often struggle to reconcile the demands of the new educational policies or reforms with their beliefs about teaching and assessment. This is a core educational problem which teachers in Singapore and elsewhere often experienced. More attention and understanding are needed if school leaders and policy-makers are serious about helping teachers to better manage and handle tension in assessment.

There may be many approaches to help teachers better manage AT. However, it will only be useful and meaningful if school leaders and policy-makers can seek first to understand teachers' distinctive experience of AT in greater depth before they could further explore more effective ways to help them. This is a core educational problem in my practice that past studies and scholarship have not been able to address adequately and satisfactorily at the moment. To be more precise, there is a lack of studies that investigate teachers' experience of AT from the perspective of the “experiencers”, although AT has been inevitable in their teaching. Teachers' unique, distinctive and rich descriptions of their experience of AT are not heard directly in past studies. Teachers' articulations regarding their experience of AT are muted. As a result, past scholars are not able to make sense of the different ways in which AT could be ‘better’ (or ‘poorer’) experienced, which I think is crucial insights for policy-makers and other stakeholders of education to understand and thus provide more concrete and effective ways to help teachers better deal with AT in practice.

In the Singapore context, teachers inevitably experience AT at the policy, programmatic, school and classroom levels when teaching and assessing students' learning from a system and culture of performativity. This is especially prominent when there is a call for assessment change to focus on AfL in an SA-dominated education system. In this research, I define AT as “teachers' feelings of struggle with competing considerations in their assessment practices”. This definition is a combination of literature reviews on AT and my experiences as a curriculum planning officer and specialist, middle manager and language teacher in school, as well as working with classroom teachers for more than 15 years. My experience and the above empirical studies informed me of another reality of how teachers experienced and struggled with policy changes such as the TLLM. However, past studies did not specifically investigate how teachers experience tension in the area of assessment change. My research could close this gap as I believe that only with benefit of empirical investigations (similar to those done by Hogan, 2011; Hogan et al., 2013; Tan, 2008b) are scholars and stakeholders of education able to gain insights on teachers' struggles regarding their experiences of AT. To this end, the context of my practice serves as an interesting and important research ground to investigate how teachers in a successful education system experience and deal with AT brought about by new challenges in assessment.

1.3 Situating CL teaching and assessment in the Singapore context

At the point of writing this thesis, I had already served as a secondary school CL classroom teacher for five years; a Mother Tongue Language (MTL) Head-of-Department (HOD) for five years; a curriculum planning officer for five years, and for the past three years, I have been serving as a senior specialist in the Curriculum Planning and Development Division (CPDD) of the MOE. While my teaching role in the school included carrying out instructional programme and assessment, the role of a HOD included setting direction and providing professional leadership for the department to ensure teaching quality and students' achievement. At CPDD, I have been developing centralised CL curriculum for secondary school learners. I had opportunities to interact with many classroom teachers and colleagues from the Singapore Examinations and Assessment Board (SEAB) and teaching training institutions such as the National Institute of Education (NIE), the Academy of Singapore Teachers (AST) and the Singapore Centre for Chinese Language (SCCL).

In those interactions I had with CL teachers, I observed that many of them often experience tension between the use of English and Chinese in their daily work. There is a naturally inherent language tension when using Chinese in a context where English is the dominate language and the *lingua franca* for different races to communicate with one another. As it is the official and common vehicle for communication in Singapore, English is also the main language of instruction in all the subjects in schools, with the exception of the MTLs.

CL teachers experienced much AT at the policy and practice level. Among all teachers in Singapore, perhaps CL teacher is the best representative of a CHC as Chinese language and culture are closely related. In many other jurisdictions alike, language teachers are seen as an information or "cultural transmitter" who do not only treat language as a means of communication, but more importantly as an end of the language study (Xiong, Li & Qu, 2015). CL teachers in Singapore are assumed to preserve and promote the language for their students to understand and develop their unique identity through a deeper appreciation of the Chinese culture (MOE, 2010).

CL as a Mother Tongue Language (MTL) is volatile due to its specificity in Singapore as its teaching and assessment are easily influenced by both international and local factors. Internationally, CL as a learning subject has drawn much global attention

in the recent decades because of the prominent increasing number of learners worldwide. Indeed, about one-fifth of the world population speak Chinese and it is the third most spoken language in the United States (Xiao, 2011). In the last few decades, education systems across the globe have moved quickly to embrace new trends of CL learning and assessment. While English is undoubtedly the global lingua franca of the current era of globalization, it is not without challenges in the recent decades. Wang (2004, 2007) has earlier proposed a notion of viewing CL in the US as the human, cultural, and social capital for individuals, groups, and the society-at-large. Elsewhere, Ding and Saunders (2006) have noted that efforts by Beijing to promote the CL and Mandarin's increasing utility as a lingua franca for trade have significantly increased the resonance of Chinese culture abroad.

Locally, the same trend is also observed in Singapore where Mandarin Chinese, has both pragmatic and non-pragmatic functions. Scholars have argued that this language should be learnt to not only take advantage of China's economic transformation, but also because it is an identity marker for Chinese Singaporeans (Stroud & Wee, 2007). Wee (2003) termed this approach in language education policy planning as "linguistic instrumentalism" (p. 211), which is a view of language that justifies its existence in a community in terms of its usefulness in achieving specific utilitarian goals, such as the access to economic development.

While both international and local factors seem to have favoured the learning of CL, the MOE recognises that the language environment at home is changing and will result in a variation of linguistic exposure and abilities (MOE, 2010). Contrary to the global increasing number of CL learners, there is a decreasing trend of Chinese primary school students using CL at home. Among ethnic Chinese students, the proportion with English as the most frequently used home language rose from 28% in 1991 to 59% in 2010 (MOE, 2010). To this end, CL teachers often lament the drop in the proficiency level of students' CL standard and are increasingly feeling much more tension in teaching and assessing their students' learning. In assessment, it was proposed that a "performance-based assessment" (a fully Information and Communication Technology- or ICT-delivered system) to be introduced to measure students' attainment of language proficiencies. This new policy requires teachers to change their current practices in assessment and inevitably invites much tension.

Levels	Tension between	
Policy	<ul style="list-style-type: none"> • The intents of MTL Policy (The Bilingual Policy) • Assessment change due to students' profile 	<ul style="list-style-type: none"> • MTL as a subject in the high-stakes national examination for placement purpose • Teachers' readiness in assessment change
Practice	<ul style="list-style-type: none"> • The traditional non-integrated language teaching and assessment • Syllabus content and depth (creativity, critical thinking skills, etc.) coverage and individual needs 	<ul style="list-style-type: none"> • The call for 21st century skills and Communication Language Ability approach that emphasises interactions, task-based and authentic teaching and assessment • Teach to performativity (doing well in national examinations) and accountability issues

Table 1.3: AT at the policy and practice levels

Table 1.3 shows how AT may be experienced by CL teachers at the policy and practice levels. First, at the policy level, both students and teachers experience tension between CL as a subject in the high-stakes national examination for placement purpose and its mission in the bilingual education policy of Singapore, as well as TSLN and TLLM. According to the report by 2010 Mother Tongue Languages Review Committee (MTLRC), bilingualism is a cornerstone of Singapore's education system. There are three broad objectives in the teaching and learning of MTL, namely for "Communication, Culture and Connection" (the 3Cs). While communication is a valuable life skill, culture should be preserved by the MTL for our students to understand and develop their unique identity through a deeper appreciation of culture, traditions, literature and history. MTL also enables our students to connect with other communities (MOE, 2010, p.13). However, MTL is also an "examinable" language subject in the national examinations, such as the PSLE and GCE 'O' Level Examination for placement purpose. These two broad intents

may contradict with each other when the policy is translated into practice because teachers may think that if they spend too much time on the 3Cs, they may not prepare their students well for the national examination. This AT is not uncommon among CL teachers as they generally find it tough to balance the learning objectives of MTL and purposes of assessment in the high-stakes examination. In addition, when there is a call for change in the current assessment practice (such as taking up a fully ICT-delivered system), policy-makers may not be aware that there will be inherent tension for teachers to change their current practice.

In practice, while the MOE has announced that “our curriculum will emphasise spoken interaction and written interaction skills in addition to the four basic skills of listening, reading, speaking and writing” (MOE, 2010, p.16), my observation is that there is still a tendency for curriculum and programme to be planned using the traditional non-integrated approach of language teaching and assessment. The main reason points to the continued usage of a non-integrated approach of testing at the national level. Indeed, assessment will ultimately pull teachers' tendency to reorient their curriculum and pedagogy towards what is going to be tested (Bernstein, 1990). I observed that the CL teaching did not go according to the call for 21st century skills and Communication Language Ability (CLA) assessment that emphasise interactions, task-based and authentic teaching and assessment described in the literature (Bachman, 2000; Bachman & Palmer, 1996).

In the classroom, I observed that there was always tension between having syllabus content and depth coverage in a given learning period on one hand and teaching the students to score well in the national examinations on the other, especially for the graduating classes. Many scholars have argued that Singapore schools need to equip and encourage teachers to adopt authentic assessment in teaching and learning to develop the students' higher-order thinking (Koh, Tan & Ng, 2012; Koh, 2011; Koh & Luke 2009). Koh and Gong (2008) have reviewed that the assessment and teaching of CL in Singapore emphasise drill and practice of basic linguistic knowledge and principles. The assessment tasks seldom engaged students in higher-order thinking, real-world problem solving, and extensive communication. There seems to be a lack of opportunity for students to be exposed to authentic language use in their learning of CL.

Indeed, many teachers are still teaching to the test and have placed a very heavy emphasis on the national high-stakes assessments. They tend to place more emphasis on drill and practice of basic knowledge and skills rather than on critical-thinking skills. This could be due to the “exam-oriented” teaching culture in the local context. In fact, Liu (2011) observes that many CL teachers used the same format in the GCE 'O' Level Examination for routine classroom assessments and term class tests which he thinks is not healthy for the language development and learning of CL in Singapore.

In line with educational reforms such as TSLN and TLLM, there were four reviews on the MTL, which took place in 1991, 1999, 2004 and 2010 respectively. The recent 2010 MTL Review Committee has identified four recommendations to further enhance MTL teaching and learning (MOE, 2010). All these educational policy initiatives and language reforms have called for a parallel change in the curriculum and advocated teaching and assessment to promote deeper understanding of knowledge, higher-order thinking skills and creativity rather than traditional rote learning such as memorising procedural and factual knowledge. The reality gets more complicated when these policies changes and reviews carry different agendas. Surely these tensions from different levels are never so simple and tidy in reality and teachers would inevitably experience them differently. I believe that these distinctive experiences deserve our attention if school leaders and policy-makers are serious about helping teachers to better manage AT in practice.

1.4 The nature of assessment tension and the research question

Tension originates from the stem of “tensio” in Latin, meaning “a stretching”. According to the World English Dictionary, tension refers to “the act of stretching or straining” or “the state of being stretched or strained”. Tension as a concept is most discussed in science (or physics).

In physics, tension is the stretching or pulling force exerted by a string, rope or similar object on another object, such as a ball. In this scenario, it may not be visible or felt but tension is virtually everywhere. There are tensions exerted both upward and downward on the string, ball, scaffold and earth simultaneously. Tensions result when the formations of the net electrostatic attractions between the particles in the object are further apart from one another. Under this circumstance, there will be a force trying

to balance this change. It is the pulling force exerted by the object, trying to restore its original form from the more compressed formation of the net electrostatic attraction. (Giancoli, 2008). It is important to note that, unlike most of the studies that reported tension in teaching and assessment as undesirable, the concept of tension in physics is “neutral” because it does not carry any negative connotation.

Based on the scientific explanation of tension in physics, there is “implicitness” in the concept of tension. The implicitness is exhibited in the multitude of pulling forces (both upward and downward) that are exerting simultaneously around the string, ball, scaffold and earth. Though we know the existence of the various tensions, it is not clear as to where they are exerting and what their relationships in the setting are. The implicitness is also exhibited by the quality of being difficult to detect, seen or observed by an external party. For example, the net electrostatic attractions between the particles in the object are not seen by naked eyes.

Teachers' experience of AT highly resembles this nature of tension in physics in that the different demands or considerations at play which teachers have to manage in their assessment practices are very much similar to the stretching or pulling forces from different directions. However, it is vital to note that, unlike this basic concept of tension in physics, AT in teaching is more complicated because it is often dynamic but yet not explicitly articulated by teachers. Additionally, teachers' emotions towards assessment must also be taken into account as it greatly influences teachers' decisions and actions in assessment.

In educational settings, Berry (2007) defines tension as “the feelings of internal turmoil experienced by teacher educators as they found themselves pulled in different directions by competing pedagogical demands in their work and the difficulties they experienced as they learned to recognize and manage these demands” (p. 119). She introduced this notion of tension as a conceptual frame and analytic tool in her studies of prospective teachers' education. This notion is a way of representing and understanding the contradicting elements and complex nature of pedagogy of teacher education. I personally think that Berry's definition of tension as “the feelings of internal turmoil” is a very negative connotation and it is rather extreme to a large extent. On the other hand, Tillema and Kremer-Haydon (2005) define tensions as “the emotional awareness of a possible conflict” (p. 204). Similar to the concept of tension in physics

explained above, this definition is more neutral and not so extreme as compared with Berry's conceptualisation. King, Patterson and Stolle (2008) went further to explain that "when something sits in tension, multiple polarities act against one another, extending, stretching and straining the object to its limits or until it is taut" (p.5). That is the reason why Talanquer, Tomanek and Novodvorsky (2007) argue studying tensions and dilemmas in teaching practice can help articulate teachers' concerns and problems. Although I grant that studying tension in teaching will allow us to better understand teachers' problems, I further assert that it will be even more helpful if school leaders and policy-makers can hear from their articulations of the experiences since teachers' own voices were not heard in the majority of the past studies on tension.

Indeed, teaching and assessment are challenging undertakings which teachers have to negotiate every day in schools. Teachers have to make many considerations before they decide on what, where, when, how and to whom the assessment should take place. In reality, AT is more complex than the concept of tension in physics simply because there are multiple ways and thus multiple outcomes for experiencing this phenomenon. This complexity of AT can be seen from the many studies that report "multi-dimensional" tensions experienced by teachers. For example, while teachers may experience tensions from the two different perspectives of reading assessment in the British Columbia system, the perspective of the students and parents further complicates the issue by disturbing teachers' pedagogy and assessment in their classrooms (Field, 1991). These different tensions at play from various stakeholders of education are similar to the multiple forces that are acting on an object. While some of these tensions may be visible, most of these tensions are not and teachers experience them in different ways.

In Singapore, there are various demands from the current assessment regime and stakeholders, educational policy and initiatives, and the call for "learner-centered" pedagogy (Hogan et.al, 2013). Together with greater parental interest and involvement in education (Khong & Ng, 2005), the mixtures of tensions in assessment are highly complex.

It is because of the nature of teachers' experience of AT that this research argues for a qualitative research method. This approach would enable me to explore in-depth into teacher's rich experience of AT. As Mason (2002) has considered:

Through qualitative research we can explore a wide array of dimensions of the social world, including the texture and weave of everyday life, the understandings, experiences and imaginings of our research participants, the ways that social processes, institutions, discourses or relationships work, and the significance of the meanings they generate. We can do all of this qualitatively by using methodologies that celebrate richness, depth, nuance, context, multidimensionality and complexity rather than being embarrassed or inconvenienced by them (p.1).

The qualitative research approach will play a key role in the investigation of CL teachers' experience of AT. I will argue in more detail in Chapter 3 that phenomenography is a suitable methodology to study teachers' experience of AT.

Indeed, there is a dearth of in-depth study on teachers' distinctive experience of AT, though there are many studies that have reported cases of AT. Most of these studies are not inclusive of teachers' experience. I would argue that the situations experienced by classroom teachers are far more complex than what previous studies have reported. While Black and Wiliam (1998a) may have looked "inside the black box of raising standards through classroom assessment", I think we have yet to explore "the black box" of teachers' experience of AT. This is the core educational problem in my context which past studies and scholarships have not been able to fully address. In the same way as Black and Wiliam (1998a) conceptualised classroom assessment in terms of "system engineering", I would observe teachers' experience of AT from an "input-process-output" structure. While the AT "inputs" are very much well discussed, the "process or how teachers experience, make sense of them and turn them into outputs" is not well understood. What are teachers' conceptions of AT? What considerations in assessment do teachers focus on? What are teachers' assessment practice and how do they carry out assessment? How do they make assessment decisions when there are multiple considerations? What are their main considerations or priority in assessment decisions? How do teachers deal with these different considerations? What are the relationships among all these distinctive experiences of AT? How are these tensions resolved, if

possible? Without much doubt, these questions are real and complex and are important enough for an in-depth research.

To help answer the above complex questions, I consolidated and summarised them into an overarching research question:

What are the qualitatively different ways in which teachers experience AT?

Assessments are not easily distinguishable and neatly separable in actual situations. School leaders and policy-makers should then recognise that assessments in practice are dynamic. Investigating teachers' experience of AT provides a dynamic perspective of understanding and conceptualising assessment which diverge from the quiescent perspectives of understanding assessment.

When teachers encounter assessment reforms in new educational policy, some are flexible to change their practice. Some will abandon their continuous efforts and struggle. Others, however, find ways to manage this tension, dilemmas, conflicts or contradictions in assessment. They would bridge the gap between the promises of sound assessment and the realities of implementation. All these experiences are unique and some teachers may have 'better' experiences than others. What are the qualitatively different ways in which teachers experience AT? What separates these differences in teachers' experiences and what are the relationships among them? All these can only be explored from teachers' real experiences and articulations. The different ways of experiencing AT by the teachers are valuable insights which various stakeholders of education would benefit from knowing as it informs them of the better or poorer ways of experiencing AT.

1.5 Significance of the research

This research is significant for several reasons. First, there is a rich body of empirical studies on the taking up of new assessment initiatives or reforms in the CHCs. However, based on the review of the literature on these (in Chapter 2), few have reported how teachers experience and manage the tension derived from the new assessment ideas from the 'West' and the traditional exam-oriented practice of the 'East'. To this end, this research combines the concepts of AT and assessment practice in the CHCs educational jurisdictions to investigate the qualitatively different ways in which CL teachers experience and manage AT in their daily teaching. In addition, by

focusing on CL as a subject discipline, this research investigates how CL as a heritage language and an examination subject that has a stake in the system, invites tension in assessment. It also informs scholars and stakeholders of education how this special group of teachers experience AT.

Second, this research contributes to the research on teachers' experience of AT. Specifically, it investigates the nature of AT from teachers' perspective. Past studies have identified tension in assessment to be negative by the researchers and the experiences from the teachers are not extensively investigated, nor is it investigated using phenomenography. In particular, this research captures teachers' collective experience of AT in terms of their distinctive differences (or variation) and their interrelatedness.

Third, this research contributes to the studies on FA, SA and sustainable assessment, as well as their complex interactions and relationships. The review of over 70 pieces of empirical work on AT (in Chapter 2) shows a lion's share of work related to the many purposes assessment has to serve in the new era of educational assessment. There are fewer studies examining the ways in which teachers experience and manage their tension in assessment. This research adds to the bulk of work on FA, SA and sustainable assessment by investigating Singapore teachers' experience of AT.

Finally, as assessment policy formulation and implementation usually do not take teachers' AT experience and ways of tension management into consideration, future assessment policies planning, initiatives, reviews or reforms may be built based on these grounded and experienced contexts of teachers' experience of AT. This research may also set the stage for a series of on-going recommendations, adjustments and even interventions around the redesigning of curriculum, teaching and assessment system and policy in Singapore. By investigating and then illuminating the AT teachers experienced, this research contributes to an in-depth understanding of teachers' real assessment experience. As a result, it addresses the enabling context and situation in which teachers can have a 'better' experience of AT. It then can suggest concrete ways to help teachers better manage the tension they face so that they would not struggle needlessly to realise the full potential of sound assessment. These insights are essential to better curriculum implementation which will eventually improve teaching and learning.

When teachers often experience AT and thus have feelings of struggle with competing considerations, decision-making in assessment practices may become difficult, inconsistent and even unprofessional. As there are many competing considerations and uncertainties during the assessment decision-making process, teachers may not know what to fall back on and thus may feel a loss in control. When this happens, decisions now depend on external factors rather than on teachers' professional judgment. From this perspective, teacher empowerment may be undermined because they have to face the problem of inconsistency, uncertainty and even compromise when making assessment decisions. As a result, a series of negative issues associated with teachers' work-related outcomes, such as low job satisfaction, poor work and organisational commitment, may have a direct impact on personal motivation, classroom improvement and school effectiveness (Lee & Nie, 2014). This problematic aspect of teachers' work has been noted in much of my experience as an educator in my context and past studies and scholarships have not been able to address adequately and satisfactorily at the moment because it is under-studied. I would further caution that a continuous failure to recognise and address AT may affect a wider application of teacher empowerment and professionalization of teaching in practice.

In addition, I would add that these definitions and arguments highlighted a real experience of phenomenon in assessment practice. While some teachers may resonate with these past experiences immediately, there are still many possible ways of experiencing tensions in reality but were very often overlooked. This research closes this research gap by exploring teachers' unique experiences of AT from their very own perspectives. This research leaves it open to let teachers voice their ways of experiencing AT. Previous scholars of tension studies have opened a lens for us to explore assessment issues. In the same way as Leong and Tan (2014) have put across regarding the consistency of applying FA theory across different sociocultural contexts, we have to accept and live with certain contradictions and view dilemmas not as obstruction for us to understand assessment issues.

As an educator, I recognise similarities between my own experiences and what other studies have reported about teachers' experiences of AT at the policy, programmatic, school, classroom and individual levels. As I shall show further in the next chapter, teachers are often in situations where they have to make tough decisions on

what types of assessment to carry out at different times in their classrooms due to a variety of factors such as introductions of new policies, reforms, initiatives or reviews. This feeling of struggle is not well-recognized as an important educational practice problem and thus is poorly understood by many people.

1.6 Research method

The main aim of this research is to provide a conceptual description and analysis of the distinctively different ways in which CL teachers experience AT in Singapore. It is to investigate how they experience and understand the roles of assessment and their tension in the current education landscape and their considerations in assessment practices. In the following chapter, I argue that previous studies on teachers' experience of AT studies have not adequately addressed teachers' experience of AT in much depth. I employ phenomenography as an approach in this research. This method enables me to describe and investigate the different ways of experiencing AT by the teachers. I am interested to explore how CL teachers conceptualise and describe the AT they encountered. In addition, I am also interested to explore the relationships within and across the different ways of experiencing AT in much depth. On one hand, phenomenographic design approach could offer a framework for me to understand teachers' conceptualisations of AT. On the other hand, it also allows me to explore how they manage AT. The justifications and details of this research methodology are further discussed in Chapter 3.

The utilization of phenomenographic research to understand the conceptualisations of AT encountered by the CL teachers as the lived experience of Singapore educators will contribute not only to the research landscape but will also inform the assessment practices of other educators, as well as how AT can be better managed.

1.7 Summary

From the above discussions on the concept of AT in general, I observe that:

- (1) it is a core educational issue in my practice and elsewhere that past scholarships have not been able to address adequately and thus it deserves greater attention;

- (2) it is inevitable as it exists in all education sectors but teachers' distinctive ways of experiencing AT are often not well recognized as an important educational practice problem and thus is poorly understood;
- (3) reconciling different competing considerations in assessment by just using the principle of "Balanced Assessment" may be too idealistic and simplistic;
- (4) in Singapore, there are tensions, inconsistency and misalignment of what the educational policy and initiative have advocated in assessment because such practice change is disruptive but there is a lack of studies to understand how teachers experience these AT.

This chapter mainly describes the core educational problem of AT common to both Singapore and elsewhere, highlighting the complexity of this phenomenon. I have also brought in my practice context, significance of this research and introduced the overarching research question along the way. The method that is employed has also been briefly mentioned. The rest of the thesis is summarised as follows:

Chapter 2 is dedicated to the literature reviews of AT. Past studies revealed that AT could arise from anywhere: from assessment itself (the purposes it needs to serve) to other external factors (such as assessment reforms or policies), from teachers themselves (paradigms of assessment) to other stakeholders of education (whom formed the historical and cultural factors). In addition, the notion of "balancing assessment" is also critiqued. The "two contrasting paradigms of assessment" is seen as binary and has not been helpful to teaching, with the historical and cultural factors of assessment in the CHCs being viewed as repellent to changes in assessment. Such reviews and critiques suggested the need to separately investigate teachers' experience of AT qualitatively.

After reviewing the methodologies employed by past studies on AT, Chapter 3 discussed the use of the phenomenographic approach to carry out this research. The unique contributions of phenomenography to study teachers' experience in AT research were highlighted. The rest of this chapter was dedicated to describing the method and its concept, the data sources and collection, as well as the data analysis. How I will assumed the role of the researcher and the issues of validity, reliability and 'bracketing' were also discussed.

Chapter 4 sets out the findings of this research. I identified three conceptions of AT (passive, responsive and propulsive) and the role of a passive executor, or a rational practitioner or a re-conceptualiser. These conceptions are logically related to each other in a hierarchical fashion. Collectively, the conceptions make up the “outcome space” that represents the full variation of the themes and critical aspects (the ‘What’ and ‘How’ aspect) experienced by the teachers in AT.

Chapter 5 reviews the main findings in relation to the research questions and past studies on AT. I recapitulate the research purpose and findings, reflect on some of the limitations, before discussing the significance and contributions of the research in four main areas: (1) Understanding AT to be positive and constructive; (2) Increasingly importance of FA in an SA-dominated educational landscape; (3) Re-looking AT in sustainable assessment; (4) Re-discovering the open tradition of Confucian education from CL teachers’ experience of AT. The implications and recommendations for policy, research and practice in schools are also discussed.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Introduction

In this chapter, I review the literature on tensions faced by teachers in assessment. Previous studies on assessment tensions, dilemmas, conflicts, contradictions and struggle in assessment tended to emphasise on how AT comes about and how teachers should reconcile this tension. Based on the extant empirical research on AT, both elsewhere and in Singapore, this chapter identifies the gaps in the current body of research that this dissertation may contribute towards filling.

2.2 Literature search method

This chapter examines both empirical research and conceptual papers on teachers' report of AT ranging from primary school to post-secondary education that includes teacher education.

As an educator, I often recognised similarities between my own observations, experiences and what other teachers reported from their experiences of AT through my conversations with them. Teachers often struggle with their many considerations in assessment as they carry out their teaching duties in schools. Bearing these experiences in mind, I started to search for literature that discussed AT, using the following keywords that have repeatedly been used by CL teachers in Singapore: "educational assessment tension", "assessment dilemmas", "assessment conflicts", "testing tension" and "struggle in assessment". These keywords are closely related to the definition of AT in this research. As defined in Chapter 1, it is "teachers' feeling of struggle with competing considerations in their assessment practices". This preliminary search was still exploratory in nature because I had to decide precisely how deep I wished to explore the topic. As the arguments found in the preliminary searched literature resonated with my own experience of AT, I knew I was heading in the right direction. Then I continued to search for more literature on AT. I employed two techniques for searching the literature: the online database approach and the ancestry approach. To canvass the literature in English, I searched Google Scholar, Education Research Complete, Academic Search Premier, Education Resources Information Center (ERIC), PsycINFO, EdResearch

Online, Hong Kong Education Bibliography Database, JSTOR, ProQuest Education Journals, Taylor Francis Online (Education Collection), Educational Assessment Evaluation and Accountability, last but not least ProQuest Dissertations and Theses database from 1989 to 2018.

I limited the search to reviewed journal articles, dissertations and theses, as well as scholarly books. From the over 200 abstracts returned, I removed those in agriculture, medical and nursing education. Then I re-reviewed the remaining to select those that had a K–12 grade or post-secondary educational and academic focus. They had to have either some sort of empirical work or theoretical literature review. I eliminated opinion papers, article or book reviews, or recommendations and guides for teachers.

The literature search resulted in 62 empirical researches (including 9 dissertations and 21 reviews or conceptual papers (total 83) that fit the purpose for this literature review based on the following inclusion criteria: (1) reported from 1989 to 2018 (29 years); (2) published in English; (3) reviewed journal articles; (4) empirical studies; and (5) involving teachers' report of tensions, dilemmas, conflicts, contradictions and struggle in assessment.

Each study was reviewed by checking and analysing their: (1) basic publication information (i.e., authors and year of publication), (2) participants (e.g., personnel, sample size, and country), (3) focus level (primary, secondary or post-secondary), (4) data collection method, (5) data analysis method, (6) types and dimension of tension (whether it is "two" or "multi" dimensions), and if there are an (7) underlying theoretical framework. The Appendix A shows a summary of the 83 studies reviewed in chronological order. A critical review is presented in the following sections, grouped into five main categories that include: (1) AT comprising different assessment purposes, (2) AT and the "two contrasting paradigms of assessment", (3) AT caused by new assessment reforms and policies, (4) Singapore teachers' experience of AT (5) AT and the history and culture of assessment in the CHCs.

I started off with a review on AT comprising different assessment purposes because this phenomenon is universal and thus is applicable to any context. Following closely is the review on the "two contrasting paradigms of assessment", which is further developed from the different assessment purposes. Then, I went on to review new assessment reforms and policies in which teachers' experience of AT is further

intensified. Next, I followed up with a closer review on how AT has been deemed to be negative and undesirable, paying special attentions to AT between new policy and practice tension, as well as teachers' beliefs and practice tension. While these four areas are still very much applicable to any context, I zoomed in to review Singapore teachers' experience of AT, paying special attentions to the concerns of assessment purpose, as well as new reforms or policies in assessment. The sequence of the review is completed with literature on AT and the history and culture of assessment, focusing mainly on how new assessment ideas affect teachers' assessment practice in the CHCs.

2.3 AT comprising different assessment purposes

The multiple duties or purposes of assessment may be in conflict with one another. Scriven (1967), Rowntree (1987), Broadfoot (1996), Boud (2000), Harlen (2006), Newton (2007), Moss and Brookhart (2009), Carless (2012) and many other scholars have attempted to discuss the aims, uses, functions, roles or purposes of assessment from a variety of perspectives. Truly, depending on the perspective one takes, there are different purposes of assessment. For example, Newton (2007) has identified 18 uses of assessment while Scriven (1967) has only pinned assessment down to two purposes. Generally, traditional educational assessments are framed toward summative purpose (for testing, grading, evaluation) while the more recent constructivist learning theories are more concerned with social and cultural constructions which promote student learning and improvement.

Besides FA and SA, assessment(s) serves other purposes. One such example is sustainable assessment, also known as the "third purpose of assessment"—assessment to foster learning throughout life (Boud & Falchikov, 2006, p.400) adds complexity to teachers' assessment experience and tension. Teachers may struggle to find an assessment that is fit for its intended multiple duties.

These multiple duties or purposes may be in conflict with one another. As teachers make assessment decisions, they will have to consider these various purposes. For example, Firestone (1998) has termed "the technical tension" (p. 176) to be the type of AT originated from the different purposes central assessments were expected to serve. In higher education, Ecclestone and Swan (1999) have conducted a research on 11 teachers from four universities and concluded that there was a growing tension

between support for selective norm-referenced assessment and the more accessible and motivating criterion-referenced assessment which could serve different purposes.

Subsequently, Reis and Villaume (2002) discussed the wide-scale use of portfolio assessment in teacher education and found that the university supervisors, cooperating teachers and former pre-service teachers experienced the inherent tension of using portfolio assessment which was designed both to foster professional learning and for certification. More recently, Harman and McDowell (2011) reported the AT over the roles of “the expert guide”, “the tutor”, “the traditional teacher”, “the objective assessor” and “the gate-keeper” by the university lecturers in the UK. These multiple roles imply that assessment has to serve different purposes and may create tension for the teachers.

To this end, for the past two decades, educators have attempted to use the notion of “balanced assessment” to fulfil different purposes of assessment so that it can enhance teaching and learning. For example, the National Education Association (2003) in the United States suggested that “To maximize student success, assessment must be seen as an instructional tool for use while learning is occurring, and as an accountability tool to determine if learning has occurred. Because both purposes are important, they must be in balance” (NEA, 2003, p.1). This idea of a balanced assessment emphasises the importance of assessment serving both “an instructional tool for learning”, as well as “an accountability tool”. This take on the balancing of the “double-importance” of assessment purpose seems to characterise different types of assessment to be opposite, separated and not having any constructive interactions.

Along the same vein, Chappuis, Commodore and Stiggins (2010, 2017) have been directing educators on how to plan an assessment system that is “balanced”. They proposed that “a balanced assessment system serves a variety of purposes, uses a variety of measures, and meets the information and decision-making needs of all assessment users at the classroom, building, and district levels” (Chappuis, Commodore & Stiggins, 2017, p.4). In my opinion, this definition of a balanced assessment system is comprehensive but too simplistic and idealistic because it assumes the complex assessment purposes needed by different users could be brought about into coherence by an overly simple notion. More importantly, it is impossible to meet the needs of all assessment users of different levels because they require different data for various

purposes at varied stages of students' learning. To this end, I think such a system is highly idealistic and would hardly be applicable in practice.

In Singapore, the first mention of “balanced assessment” started in 2013 when the MOE initiated a move for schools to shift “Towards Learner-Centred and Balanced Assessment”. This principle is the underpinning philosophy for assessment change in its assessment document for secondary schools released in 2013. The notion of “balanced assessment” may be traced back to 2009 when the Singapore Examinations and Assessment Board (SEAB) Chief Executive, Ms L. C. Tan suggested that there is a need to “audit our practices for balance”. She proposed:

The key word here is ‘balance’; neither excess nor neglect. Our journey is not about having examinations or no examinations – rather it is a search for a healthy mix of the two types of assessment. How do we balance the use of examinations for placement with use of assessment to inform learning? (Tan, 2009, p.2)

It is important to take note that the Chief Executive highlighted “balance” as the key, which means it assumes a pivotal role in ensuring sound assessment practice. However, the notion of “balancing assessment” is not easy in practice because “balance” carries a high degree of subjectivity and can be subjected to many interpretations. For examples, teachers may try to balance their assessment between FA and SA; they may also focus on adjusting their assessment practice between giving students quantitative and qualitative feedback. They may also shift their assessment from “standardised” ones to more “personalised” ones in order to “balance” their assessment practice. There are many possibilities in conceptualising “balanced assessment”. Further, there is no criteria or standard to show that any assessment system is “balanced”.

Additionally, the phrases “neither excess nor neglect” and “a healthy mix of the two types of assessment” are extremely ambiguous and did not offer any concrete and constructive instruction on how educators at different levels (policy, programmatic, school, classroom and individual) should and could use assessment in a “balanced” fashion. It is not clear as to how exactly teachers should conceptualise and operationalise “balanced assessment” in schools and subsequently to resolve this

problematic aspect of teachers' work. Indeed, assessment could be defined in several ways depending on one's perspective. Often, teachers encounter conflicts and tension when carrying out assessment that serves more than one purpose (for e.g, "double duty" (Boud, 2000)) and finds it hard to "balance" in practice. Having "a wrong balance" or being "not in balance" implies that teachers are not able to ensure that their assessment is "fit for purpose" because their original intended purpose may have to give way to achieve "balanced-assessment".

The above conceptualisations of "balanced assessment" suggest that this notion is not so simple as to merely place an equivalence between the two types of assessment (examinations for placement and assessment to inform learning). It involves many dimensions and requires all stakeholders coming together, with a common understanding, working toward a common vision. Without a common understanding at all levels, teachers will have many assumptions about assessment and be unsure if they have achieved the "balanced assessment". To this end, the notion of "balanced assessment" could be said to have totally ignored the tensions in assessment possibly experienced by teachers in practice.

More recently, Leong and Tan (2014) have reviewed the various education reforms and traced how FA was conceptualised and played out by Singapore schools. They observed the challenge of:

How teachers can reconcile a possible dilemma between helping students achieve grades in tests and examinations for public accountability, and also make any new classroom assessment 'count' for student-centric aspirations, may remain a matter of policy and practical tension (Leong & Tan, 2014, p.604)

They then went on to suggest ways to overcome this AT by proposing to address the problem of balancing "AfL" and "AoL". It is exactly this notion of "balanced assessment" which has created AT for teachers. For this reason, any suggestions to "re-balance" assessment will suffer from the same limitations of "balanced assessment" discussed above.

Another issue with a simplistic notion of “balancing assessment” is that it fails to take historical and cultural factors into consideration. The notion of “balanced assessment” assumes that these factors will not affect how teachers carry out assessment and thus will be able to apply this notion in assessment. I will show later that these factors play a vital role in teachers’ experiences of AT.

The above past studies showed that to “balance” assessments and to ensure that they are fit for their intended purposes is indeed challenging and difficult to achieve. The notion of “balanced assessment” was conceptualised with the intent to fulfil the different purposes required by different stakeholders of education. However, this notion in reality has not been helpful. Teachers trying to “balance” their assessment practice would still experience much tension. In fact, the notion of “balanced assessment” totally ignores AT experienced by teachers. To this end, I argue that by investigating teachers’ experience of AT, it provides stakeholders of education with an alternative lens on how the multiple purposes of assessment required of them could be better understood. Subsequently, stakeholders of education would also have a better understanding of the enabling conditions in which ‘better’ experience of AT may be achieved.

2.4 AT and “two contrasting paradigms of assessment”

While assessment may have to serve multiple duties or purposes in reality, it is not uncommon to encounter instances where there is an over-simplification of assessment as being merely “two contrasting paradigms”. The “two contrasting paradigms” always appear in “pairs” of various dimensions in assessment.

Many writers have argued that depending on the perspectives, there are different aims, uses, functions, roles or purposes of assessment, which may be in conflict with each other. However, educators’ assessment discourse and practice seemed to have oversimplified the multiple purposes of assessment into just a binary concept of “two contrasting paradigms of assessment”. This view has been influential in that many past studies have identified AT when assessment has to serve the distinct binary purposes (for examples, “assessment for improvement versus assessment for accountability”; “FA versus SA”; “AfL versus AoL” and so on). If there is always only “a

pair” of contrasting paradigms in assessment, it implies a prevalence of tension between these two opposite ideals because they will always need to compete with each other.

According to Shepard (2000, 2001), the introduction of new theories of curriculum and learning has given a new role to assessment. For example, the constructivist-oriented assessment requires new pedagogical practices and clear expectations of desired learning outcomes for both teachers and students. It is formative in nature and a theoretical framework of FA has been offered by Black and Wiliam (2009). Another good example is the slogan that “all students can learn”, which is intended to refute the traditional conceptions that only the elite students could learn and excel. Conversely, Keogh (1980) explained that:

The behaviouristic orientation has been felt keenly in the area of assessment, where in the past heavy reliance was placed on the use of standardized instruments administered in an office and leading to classification (or failure to classify) and only vague prescriptive statements.

In this case, the contrasting pair of “constructivist” and “behaviourist” orientation seem like a binary system in assessment thinking. This binary idea is a pair of assessment concepts with directly opposite meanings or purposes. Similar to the concept of “either on or off” and “either black or white”, the binary characteristic of the “two contrasting paradigms of assessment” is that it is always a “zero-sum game” and this creates tension for teachers. For example, if teachers have the thinking that “FA's gain is equivalent to SA's loss” or “time spent on FA is time that is taken away from SA”, then there is a possibility that their assessment practice will be filled with a lot of tension. To this end, to talk and learn about these “two contrasting paradigms of assessment” allow a better understanding of the considerations that are going through teachers’ minds, as well as the tension they experience. For example, Black (2003) argued that unless the system of SA for accountability and certification of students is fundamentally reformed, the great potential of classroom FA for improvement in standard cannot be fully realised. His statement implies there are two competing and contrasting ideas in assessment that is binary. In order for FA to work, SA has to give way. His statement also hints of tension

for teachers with the current assessment system favouring public accountability. Ewell (2009) captured and summarised the different “pairs” of contrasting paradigms of assessment in the table below, showing how assessment can be binary and contrasting in many dimensions:

Assessment for Improvement Paradigm		Assessment for Accountability Paradigm
Strategic Dimensions		
Intent	Formative (Improvement)	Summative (Judgment)
Stance	Internal	External
Predominant Ethos	Engagement	Compliance
Application Choices		
Instrumentation	Multiple/Triangulation	Standardized
Nature of Evidence	Quantitative and Qualitative	Quantitative
Reference Points	Over Time, Comparative, Established Goal	Comparative or Fixed Standard
Communication of Results	Multiple Internal Channels and Media	Public Communication
Uses of Results	Multiple Feedback Loops	Reporting

Table 2.4: Two Contrasting Paradigms of Assessment (Ewell, 2009, p.8)

According to Ewell (2009), assessment could be grouped into two contrasting paradigms (either for improvement or accountability) that can be further divided into strategic dimensions (intent, stance and ethos) and application choices (instrumentation, evidence, reference points, communication and use of results). The typical example is “the pair” of FA and SA. Similar to the above-mentioned conceptualisations of “balanced assessment”, these different “pairs” of contrasting paradigms of assessment ignores AT possibly experienced by teachers and the complexity and interactions in which

assessment concepts are experienced in reality. For example, this view again ignores the “third purpose of assessment” (sustainable assessment) and many other new assessment concepts such as “Assessment as learning”, “Learning-oriented assessment”, “Dynamic assessment” which may have overlapping or additional roles in assessment. Each of these assessment carries certain characteristics and assumes different roles. The tidy perspective of such “two contrasting paradigms of assessment” will remain inadequate if stakeholders of education do not analyse the practical and actual contexts of those who are experiencing the assessment concepts. Indeed, Bennet (2011) argues that FA “does not yet represent a well-defined set of artefacts or practices. It is both conceptually and practically still a work-in-progress” (p.19).

Despite Bennet’s (2011) argument, many writers continue to synthesize assessment realities into this neat conceptual semantics which are not able to reveal and encompass the complex experiences encountered by teachers. For example, Bishop et al. (1999) studied users’ perception of the General Certificate of Secondary Education (GCSE) and found that teachers experienced tension with the addressing formative objectives (for improvement) of the GCSE while it was essentially a SA in nature. In the same vein, Bonner and Chen’s (2009) large-scale study on the 222 teacher candidates’ experience of the tension between “constructivist learning theory” (FA) and “principles of educational measurement” (SA).

The above reviews showed that merely attributing AT to different “pairs” of the “two contrasting paradigms of assessment” is not helpful and unproductive because assessment in reality is not binary and has to assume many and sometimes contradicting duties, which are not truly captured by just tidy “pairs”. More importantly, these contrasting “pairs” of assessment also imply a “zero-sum game” relationship between the binary of assessment. It means that if one purpose in assessment is served, the other purpose will have to give way. Such view also assumes that there will not be a “win-win” situation where both purposes in assessment could be served. Consequently, this opposing relationship of assessment will limit our lens on how different types assessment may work together to achieve productive learning.

Similar to the notion of “balancing assessment”, the “two contrasting paradigms of assessment” also fails to take historical and cultural factors into consideration. I wish

to show in the later section that these factors play a vital role in assessment practice and tensions are often reported when teachers are told to change their practices.

2.5 AT caused by new assessment reforms and policies

While the AT reported in the above section were on-going struggles of teachers' experience of assessment in their daily teaching, new assessment reforms and policies have intensified this tension and teachers found it to be even more challenging in dealing with this phenomenon.

Many studies have reported how educational and testing reforms had created tensions, stress, conflicts, contradictions or dilemma for teachers in practice (for e.g Black, 2005; Firestone, 1998; Kim et al., 2013; McKay & Brindley, 2007; Tan & Leong, 2014). These studies found that teachers experience AT because the new educational reforms or policies required them to make changes to their current practices. The agendas behind these policies or reforms may be against the will and beliefs of the teachers.

When teachers encountered tension in assessment, it is reported to be problematic and challenging. In these instances, teachers have to juggle, negotiate and mediate the competing demands by various agents. They have to spend time and effort to respond, manage or reconcile the AT.

According to Fullan (1982, 1986), educational innovation, reform or change in practice can be defined in relation to three aspects. As new agendas are found in the new educational policies or reforms, they may diverge from the original classroom teaching and assessment that teachers used to practise.

Typically, educational reforms in the United States, United Kingdom, Australia and many other countries are moving toward providing directions for teaching and learning through defining and showing standards for both teachers and students (Evan, 2002; Hamilton, Stecher & Yuan, 2012; Lindsay & Lewis, 2003; OECD, 2011; Zgaga, 2013). These standards, such as teaching competencies for teachers and language and mathematics literacy for students, are the main focus of these reforms and would play a vital role in assessment (Delandshere, 1996). In these studies, standardised tests have been a key element in accounting, monitoring and evaluating the performance of students, teachers and schools. For example, both Black (1995) and Troman (1989) have

reviewed the 1987/8 central government initiatives on a National Curriculum, and National Assessment and Testing in the United Kingdom. They expressed concerns over the different standards and purpose of external tests and teachers' classroom assessments.

Paechter (1995) investigated teachers' dilemma in carrying out the statutory assessment tasks of Design and Technology (D&T) curriculum under the "Standard Assessment Task" (SAT) in 1989. Under this policy, teachers were required to assume the different role from their normal instructional lessons. This AT between good teaching and "fair" assessment was imminent during the actual execution of the practical tasks. Further, Paechter (1995) also reported that teachers' responses to the assessment dilemma in the D&T curriculum took different directions from the intended policy. While some followed strictly to the test regulations, others tried to interpret the regulations in such a way that their actions at least appeared to be within the rules. Some made their own decision based on their thinking of teaching and assessment.

Indeed, the majority of the empirical studies confirmed that new educational policies and reforms would require teachers to change their teaching and assessment practices when faced with new demands for improvement (Ruiz-Primo, 2006). These responses and changes were diverse and thus could happen in many ways. This diversity means that individual teacher's experience of AT is distinctive. However, past studies have not investigated these collective experiences and their relationships among these ways of experiencing AT.

How teachers consider, interpret and translate those policies and reforms from the institutional level to the school or classroom level directly affect the quality of their experience in assessment. Past studies have shown that there are diverse ways of responding to new educational policies and they may be highly unpredictable.

Lefstein (2008) found that some teachers may accept the new educational policies or reforms and change their assessment practices accordingly. Some may pay "lip-service" to agree with the change but still stick to their old practices. Others may just ignore the policies or reforms and carry on with their existing practices. On the other hand, some may openly criticise and show their unhappiness with policies or reforms, they never make any changes. A handful may become flexible enough to make changes to suit their current practice under the new initiatives. Of course, there are also teachers

who think that they have followed the new policy closely but their classroom practices remain the same at “drill-and-practice” (Cohen, 1990). Indeed, teachers’ varied interpretations and thus the enactment of new policy or reform depends on the teachers’ choices. Teachers may utilise their limited autonomy by fully accepting new educational policies or reforms without questions. Alternatively, they may also question, repel, reinterpret and undermine them. Therefore, there is a direct relationship between the interpretations of the policy or reform and their assessment practices and experiences (Bowe, Ball & Gold, 1992; Helsby, 1995).

My investigation is not limited to how teachers experienced AT specifically caused by new educational policy or reform. I value their unique experience of struggles with multiple competing considerations with assessment decisions and practices. I also value their considerations and priorities when interpreting new educational policy or reform. More importantly, these different considerations add insights to how teachers are experiencing AT, which is valuable for policy-makers and other stakeholders of education in their future planning of assessment reforms or policies.

Teachers experienced a certain degree of conflict and tension when they responded to new educational policies or reforms (Cohen & Ball, 1990). Indeed, Vandeyar (2005), quoting earlier works done by Black (1993), Crooks (1998); Blok, Otter & Roeleveld (2002) and others, reported that teachers often struggle to reconcile the demands of the new educational policies or reforms with their beliefs about teaching and assessment. This is a problematic aspect of teachers’ work because teachers must handle multiple competing considerations in assessment decisions.

Many other studies have reported a similar practice problem. In Australia, Kostogriz and Doecke (2011), Dreher (2012) and Hardy (2013) explored how the new national testing policy, the National Assessment Program for Literacy and Numeracy (NAPLAN), has influenced teachers’ practice in teaching and assessment. Dreher’s (2012) study found that teachers were experiencing tension between adherence to the testing policy and responding to the individual learning needs of students. Hardy’s (2013) study also revealed that teachers were caught in between “centripetal forces” and “centrifugal force” of teaching. The former favours standardisation of practices but the latter prefers to dismantle such an approach. Kostogriz and Doecke (2011), on the other hand, found that teachers were grappling with the tensions between state-wide testing policy and a

sense of their professional responsibility for their students. Arkoudis and O'Loughlin (2004) has earlier discovered teachers' AT between the "administrative purposes" (for accountability) and "educational purposes (for students' language development) under the English curriculum and standards framework (CSF) 2 in Victoria. These cases shared similarities with the "two contrasting paradigms of assessment" discussed in last section where I pointed out that the binary idea has not been helpful to teaching.

In addition, there were some studies that explored the experience of higher-education teachers' tension in assessment. Volkman et al. (2005) investigated the undergraduate physics course teachers' experience of the inquiry-based science instruction policy and found that on one hand, the lecturers felt it was important for the test to assess conceptual understanding. On the other hand, they found problems as the final test did not measure students' use of inquiry to develop evidence-based explanations.

In summary, these studies reported AT caused by new assessment reforms and policies which seem to be undesirable because AT can have a negative impact on teaching and learning.

2.6 AT as negative and undesirable

As explained in previous chapter, although the concept of tension in physics is "neutral", it carries certain negative connotation in many people's perception. Many studies confirmed that such connotation did not only appear at the perceptual level. For example, Harlen and Deakin Crick (2002) concluded that there was international evidence that the tensions between AfL and AoL are very real. They further posited that if such AT is not reconciled, it could have a negative impact on learning and teaching. In addition, other studies confirmed that AT is undesirable because it added on to teachers' workload and roles. It also affected their well-being in the mental, emotional and psychological domains.

For the past few decades, studies in the United States, England, Australia and elsewhere reported the negative and undesirable effects brought about by the introduction of new assessment policies. For example, Firestone (1998) reported AT generated from the new assessment policy in England and Vermont in the 1990s. Specifically, the study identified two types of tension (the technical and political tensions

of assessment). It was found that the new assessment system “created manageability problems, drastically increasing teachers' workload and decreasing time for instruction” (Firestone, 1998, p. 180). In fact, some teachers rejected the new assessment policy as they felt being oppressed due to the failure of the government to provide timely materials to help teachers prepare students for the test.

In the United Kingdom, Paechter (1995) inquired into the effects of the new SAT policy on teachers' assessment practice in the D&T. He discovered teachers' frustrations over the dilemmas faced by them with regard to their conflicting roles as assessors and instructors under the policy. Gioka (2008) used interpretive approach to study how nine science teachers taught GCSE and A-level science classes coped with tensions between FA and SA duties. It was found that it was difficult for teachers to fulfil their dual and challenging duties. Teachers needed support to assume these double duties. The study further concluded that the external SA dominated and distorted the teaching and assessment of science coursework.

Hammerness (2004) also identified a high school science teacher's experience of AT. It was between her personal ideal thinking in assessment and the standardized testing initiative under the Massachusetts Comprehensive Assessment System (MCAS). Hammerness (2004) further concluded that “The state policies intrude upon practice, and led her to unsettling concerns, fears and doubts. The dramatic opposition between her vision and that of the state leads her to questions about her vision, the abilities of her students, as well as her own ability to teach.” (p.37)

In language assessment, McKay and Brindley (2007) investigated the changing role of classroom assessment in the context of the Australian Language Levels Project, a national reform for English as Second Language (ESL) and Languages Other than English that promoted classroom assessment. The study identified problems arise from the tensions between policy and practice. It was further found that both teachers and students are finding it difficult to deal with the extra pressures caused by external demands.

Many studies also reported the negative effects of AT between teachers' beliefs and their assessment practices. Field (1991) examined the assessment beliefs and orientations of 17 educators and the circumstances that shaped the interpretation of a reading assessment in the elementary public-school system in Canada. The study

revealed that the reading assessment was filled with tensions, contradictions, and dilemmas that could not be solved or eliminated simply through application of technical rational thinking (p.211). Using mixed-method, McMillan and Nash (2000) investigated 200 mathematics and English teachers' beliefs and practices in classroom assessment in the United States. It was found that teachers are constantly fighting to reach a balance between their beliefs about education and the realities of the classroom environment and external demand of assessment.

James and Pedder's (2006) survey of 558 teachers in England found that there were AT between promoting learning autonomy and performance orientation, which was considered to be values-practice gaps. Over half of the participants reported that they are not able to sustain classroom assessment practices in line with their values. Some of these constraints were beyond the control of the teachers.

Deng and Carless (2010) used four case studies of teachers in two primary schools in China to study the impact of the introduction of the English communicative language teaching and assessment. Many factors were found to be responsible for impacting on the interplay between communicative language approach and examination preparation. It was reported that teachers were experiencing some pressure from the national examinations when they were using the communicative language approach in their teaching. Another important finding was that teachers' beliefs about language assessment, rather than the public examinations, acted as a constraint to the implementation new language teaching approach.

The above studies have shown AT to be problematic because there was a gap between teachers' current assessment practice and the newly introduced assessment policy. In some instances, there was also a gap between teachers' belief about assessment and the required practice they must assume in practice due to accountability. To this end, AT affected teachers' workload and could be detrimental to teachers' well-being in the mental, emotional and psychological domains. In short, AT is usually reported and construed in past studies to be negative and undesirable to teaching and learning. According to my observations and interactions with CL teachers in Singapore, they too suffered from such values-practice gaps in assessment. For example, CL teachers are commonly required to assume the role of cultural guardians who are supposed to infused Chinese culture and values into their daily language

teaching. However, it was not uncommon to hear CL teachers unwittingly complain about their role to teach Chinese culture and values (including moral education) in practice because they are more concerned about their students' language learning, especially their performance in the national high-stakes examination. To a large extent, this AT of values-practice gaps is peculiar to CL teachers because other subject teachers are not required to officially impart Chinese culture and values.

So far, both in the 'West' and the 'East', there has not been any study on new assessment policy or reform that was reported tension-free. Likewise, in Singapore, assessment reforms have also met with similar challenges and problems in the recent decades which have contributed to the insights of AT research.

2.7 Singapore teachers' experience of AT

As mentioned in the previous section, there has not been any educational reform or policy that was reported "tension-free". Singapore has been implementing educational reforms which mainly originated from the 'West' for the past decades. However, it is also considered as a contemporary CHC that has deeply-rooted "exam-oriented" education history and culture. Singapore may be described as a cosmopolitan city that embraces cultural diversity from both the East and West. To this end, it may serve as an important research site to explore how these two cultures in education interact with each other.

The following four studies reported situations where teachers have to change their old teaching and assessment practices due to new assessment initiatives or reforms. Within the limited research conducted in Singapore on assessment, Lee (2009) has used the Cultural-historical activity theory (CHAT) as a lens to explore how a student experienced the Knowledge Building (KB) reforms that required a change in pedagogy and assessment practice in Singapore. The findings highlighted some of the tensions in implementing student-centred pedagogies and assessment. Specifically, students still favoured succeeding in examinations and teacher-fronted instruction, over seeking new knowledge from application which brought about much tension to the teachers. This can be partly explained in terms of a conflict between East Asian culture manifest in Singapore (some scholars also label Singapore as a CHC) and a new and different way of teaching, learning and assessment from the West.

Norashikin et al. (2010) conducted a small scale, mixed-methods study that investigated upper primary teachers' perceptions of the PSLE English oral assessment. Teachers generally identify various challenges to oral assessment and think that there is a need to review and modify the current practice of assessing students' communication skills. To be more precise, the study found that teachers identified different challenges in carrying out this assessment task and there are limited common perceptions among them regarding the validity of the PSLE English oral assessment. As they only focused on the oral component of a large-scale, high-stakes summative testing of the EL, it is not known if such specificity were convincing enough to make a more concrete conclusion regarding teachers' thinking and experience of English Language (EL) assessment in Singapore. This study has not focused in-depth on how teachers experience assessment which I would argue to be an important insight if we are serious about helping them to better deal with AT.

Tan's (2013) study has a very different focus. Tan (2013) reported the results of a phenomenographic study on 13 primary school teachers' (teaching Science, English and Mother Tongue subjects) experiences of alternative assessment in Singapore. Conservative, pragmatic and progressive were the three conceptions of alternative assessment identified in his study by the teachers. Each conception depicted a particular manner of understanding and their use of alternative assessment in schools. The study was done at the time when the assessment system in Singapore has given much attention to alternative assessment as a means to alleviate negative backwash effects of high-stakes national examinations. From his study, Tan argued that sustainable assessment (Boud, 2000) practices were real in Singapore. Tan's descriptions on how teachers experience and understand alternative assessment was very rich. The distinctions among the three conceptions were clear. While Tan's study has indicated that there were three conceptions for teachers toward alternative assessment, it was not clear, however, that this conclusion could be applied to secondary school teachers and/or for a specific subject when alternative assessment carried a particular meaning to them. His study has also not paid special attention to AT experienced by teachers in Singapore although this phenomenon has been very prevalent in its educational system for many years. By focusing on a subject-specified study, it adds on to our knowledge of how AT is experienced by different teachers.

Kim et al. (2013), using narratives and questionnaires, together with reflective writing and focus-group discussions, found much tension experienced by teachers between the focus on new science syllabus and current assessment in their understandings of inquiry. For example, while teachers agreed with the new initiative, they also voiced that students would still need to take examinations at the end of the day. As such, on one hand, they were trying to get students curious via the inquiry syllabus together with relevant pedagogy during science lessons. On the other hand, they were also trying to get them well-prepared for the examinations. They did not see alignment between these two endeavours and thus struggled to make assessment decisions. The mixed-methods approach provided a rich description of primary science teachers' thinking about the inquiry syllabus, pedagogy and assessment. However, what is lacking was the differences or variation in the ways of how teachers experienced the new syllabus and assessment, particularly the reported AT. It will be meaningful if the study can investigate the distinctively different ways that AT is being experienced by teachers. Unless these unique experiences are captured, the policy-makers and other stakeholders of education may not be able to learn and thus suggest better ways to help teachers manage AT.

All in all, the above Singapore researches on assessment showed that teachers were not used to the new teaching and assessment initiatives and they struggled with these new experiences. Some also showed that the new demands in assessment practice somehow affected teachers' current practices and how they made professional decisions in assessment. These past studies demonstrated the complexity of teachers' experience of assessment, and clearly highlighted the necessity of further research, particularly an in-depth and rich investigation of teachers' distinctive experiences of AT within a CHC context. These distinctive and unique experiences will provide readers with valuable insights on how educational policies, reforms and reviews can be positioned so that AT is better managed in the future.

2.8 AT and the history and culture of assessment in the CHCs

As introduced in Chapter 1, CHCs are more well-known for their "exam-oriented" teaching culture as they usually favour a teacher-centred pedagogy and emphasise more

on academic performance in high-stake examinations (Luk-Fong & Brennan, 2010). This inclination has its historical and cultural roots.

The history and cultural factors related to any educational system are crucial aspects for understanding its assessment and testing (Black, 1998). CHC as an explanation has exerted an impact on research in educational psychology and assessment (Kennedy, 2016). Indeed, after almost 2500 years, Confucian thinking continues to have a great impact on contemporary Asian living, as well as on approaches to teaching, learning and assessment. Some studies have even positioned it as a barrier to assessment change and innovation (Carless, 2012; Davison & Hamp-Lyons, 2010; Hamp-Lyons & Davison 2010). To this end, the effort to continue exploring how assessment ideas from the 'West' are taken up or experienced by the CHCs will be valuable. This is because the international education landscape will continue to be dynamic and fluid in the coming decades and these insights would allow both policy-makers and educators to make predictions on the assessment practice trends, as well as the taking up of new assessment ideas from the 'West'.

Past studies found that given the historical background of the CHCs that include Singapore, the introduction of FA (or AfL) has met with massive challenges as their assessment cultures were repellent to changes. Education in these CHCs is very examination-oriented and thus may not agree with assessment that emphasise formative purpose.

Carless (2012) posited that the way examinations and their preparation were approached was most affected by the country's historical and cultural factors. After reviewing the long history of the Chinese imperial testing system as a means of social competition and control, together with his findings from the Primary English Assessment for Learning (PEAL) project, Careless (2012) concluded that in the CHCs, the strong willpower and motivation to do well in the examinations at all levels, was a highly valued characteristic of students. That is the reason why top performers in these examinations are always well-respected. As the old Chinese saying goes “十年寒窗无人问，一举成名天下知”， which means nobody ever notices the ten years of hard study one endures, but the fame will fill the whole world once honours are won in the examination. This maxim implies an education which is very examination-oriented and the emphasis is

given to the final SA, almost neglecting the formative ones. In such systems, the performance in these high-stake examinations and education is inextricably linked.

Indeed, drawing heavily from both English and Chinese literatures, my study of Top Scholars (状元) in the Chinese Civil Examinations (科举) of Song Dynasty (960-1279) also found that candidates' performance in these Confucian-oriented examinations was a "life-and-death" issue and it often implicated their families, clans and lineages. I further concluded that the Chinese Civil Examinations played a vital role in the high degree of social mobility in Song Dynasty (Chan, 2010). The Examinations provided a platform for all walks of people to move up the ladder of success in life. These historical studies or reviews informed us that CHCs (which include China, Hong Kong, Japan, Singapore, South Korea, and Taiwan) are very examination-driven (Biggs, 1996). They also favoured teacher-centred pedagogy, emphasised on academic performance in high-stake examinations, with teachers taking a big responsibility of students' academic results (Luk-Fong & Brennan, 2010). Some tensions will be created due to cross-cultural differences when the new theories of learning and assessment enter these systems. Teachers may find mismatches between these aspects of assessment practices and their "traditional" learning cultures (i.e., an examination-oriented education). As introduced earlier by Paul Black and his team members, the four effective practices or aspects of FA are questioning, feedback, peer and self-assessment and the formative use of summative tests (Black & Wiliam, 1998a; Black et al., 2003). These are situated in the examination-oriented cultures where schools focus on how to help students perform well in examinations. These cultures constantly emphasise memorising of knowledge and hard work and are highly competitive because attention is often given to the top students (Carless, 2012; Han & Yang, 2001).

The following four studies (three conducted in Hong Kong and one in Vietnam) further support my above argument. Hong Kong's historical and political background is very similar to Singapore. It was a British colonised region until 1997 with the majority Chinese speaking Cantonese. It is a typical 'East meets West' globalised city where the education systems are both inherited from the British, as well as the indigenous Chinese culture. Balancing the Chinese (CHC) and Western culture in the Hong Kong education system has been challenging. Cheng (1990) observed that there are constant struggles

to reconcile the Western values inherited from the British system and the traditional culture and values held by the local parents and employees.

Both Davison (2004) and Luk-Fong and Brennan (2010) reported teachers' experience in the assessment movement in Hong Kong from high-stake public examinations to AfL. One good example was the implementation of school-based assessment (SBA) where teachers must assume multiple roles in assessment. Both FA and SA are required of teachers. The tension evolved from the fact that teachers had more work than simply preparing students for public examinations. They also needed to be familiar with the criteria used in the SBA. Luk-Fong and Brennan (2010) reported that the change "is a concern for students, parents, and teachers alike because the final grade is likely to have an important impact on university admission and, ultimately, employment in highly paid jobs" (p. 131). This statement clearly demonstrated the examination-oriented thinking in this CHC, which may not be easily changed. It was also reported that although teachers in Hong Kong have experienced tension from the educational reforms, they were still trying their best to help students learn. They have tried to overcome these difficulties with their "personal power" and "resourcefulness" (p. 149).

Yung's (2001) case study on three Hong Kong biology teachers' perceptions of the assessment reform (i.e., the Teacher Assessment Scheme as a case) found that there are three distinct conceptions held by them which represent the different experiences and responses to the new assessment reform. Policemen, examiner and students' companion are the three metaphors used by the teachers to depict their experience. It was further found that when a teacher performs the dual task of both classroom instructor and assessor in the assessment reform, they would encounter tension.

While similar to Hong Kong, Vietnam was colonised by the French at the end of 19th century. Since then, the traditional Confucian-oriented education, was replaced by the "French-Vietnamese" education. Still, the characteristics of the CHCs are highly influential in the country. Thanh-Pham and Renshaw (2014) found that FA met with various barriers and teachers experienced tension in classroom assessment when it was introduced in Vietnam. For example, peer assessment was not favoured because in a CHC classroom, teachers (and not students) are seen as the professionals who have the authority, experience and trustworthiness to give feedback to students. As such, the

Vietnamese teachers could not implement FA as planned. In response to the AT experienced in their classrooms, the Vietnamese teachers developed FA that could at the same time formally evaluate students' learning. For example, instead of only planning the traditional end-of-term paper and pen examination, the teachers injected a mid-term group oral presentation to check students' learning progress. Many more FA such as feedback practices were built in during the process of preparing for the group presentation. As such, the transformed pedagogy or assessment practice helped students' learning at the end of the day (Thanh-Pham & Renshaw, 2014).

The above studies showed that there are conflicts and tensions between the newly introduced FA and the traditional "examination-oriented education" mostly observed in the CHCs. Carless (2012) then proposed "formative use of summative testing" (FUST) as more suitable to carry out assessment in the CHC-set schools. However, to a large extent, by treating FA and SA in the CHCs as a continuum, teachers may still require performing the "balancing" work. It may still be problematic for teachers as they must decide "when" and "how much" of their teaching to employ "what" types of assessment. Again, it is open to multiple interpretations and each teacher may develop their own versions of "balanced assessment". Confusion may result if there is no further unpacking of the notion to guide them. At this point, CL teachers in Singapore deserve special attention again for two important reasons. First, besides preparing students for high-stakes national examination in the classroom, most of the CL teachers are also required to go for national examination marking assigned by SEAB. In other words, they are assuming two official duties as a CL teacher in terms of assessment: one at the school level where assessment is carried out during the instructional process (or at the end of a school term) to provide feedback (or giving information on students' mastery of content) to adjust ongoing teaching and learning. However, CL teachers also need to change their school-teacher identity and become a rater in the national examination. Some teachers may experience conflict from these multiple roles they have to assume as a CL teacher.

Over the past decade, there have been some attempts at different levels to reconcile the tension brought about by new assessment initiatives. However, it is not known how teachers in a CHC, such as Singapore, experience and manage this AT. Singapore has a unique and idiosyncratic culture; it is also often considered to share

certain cultural similarities with other CHCs. Within Singapore teachers, perhaps CL teachers are ideal representatives of the CHCs due to their direct access to the language, culture, history and heritage of the Chinese. Further, according to Nguyen, Terlouw and Pilot (2006), Chinese teachers are strongly influenced by CHCs and they face dilemmas and tensions in the process of education borrowing from the 'West'. More importantly, CL teachers in Singapore inevitably face the AT brought about by the global trends in CL teaching and assessment, as well as its internal education and assessment reforms in the recent years. There is inherent AT in teaching the language in Singapore and much of those AT experienced by teachers reside in CL assessment.

Indeed, investigating CL teachers' experience of AT in the CHCs is significant for a few good reasons. First, one of the three main objectives in the teaching and learning of MTL in Singapore is for "culture". The MOE believes that culture should be preserved by the MTL for the students to understand and develop their unique identity through a deeper appreciation of culture, traditions, literature and history (MOE, 2010, p.13). To this end, CL teachers could be deemed to be the Chinese cultural guardians and transmitters and thus no other subject teacher could be a better representative to be the "spokesmen" of the tension experienced when it comes to how the 'Western' assessment ideas enter the CHCs.

Second, when the newly introduced assessment ideas (such as FA and AfL) from the 'West' enter the CHCs, it created a cultural tension in assessment. Naturally, the traditional "examination-oriented" and "teacher-centred" teaching culture of the CHCs was being challenged. To this end, the identity of Singapore CL curriculum, pedagogy and assessment may also be said to have been confronted.

Third, when new educational ideas from the 'West' interact with CL teachers' teaching and assessment practice, the AT they experienced would also represent the extent in which their conceptions of teaching and assessment could be compromised.

All the above would be important to cultural studies in education, especially cross- or inter-cultural research. For example, it can shed light on cultural frameworks such as the Developmental Model of Intercultural Sensitivity (DMIS) created by Bennett (1986, 2004). It is usually used to explain how people experience and engage cultural differences. The DMIS is represented in a continuum that starts from the experience of one's own culture to the experience of one's own and other cultures (Bennett, 1986,

2004). In particular, my research would be able to better understand how these “invisible” cultural differences of the ‘West’ and ‘East’ in educational assessment have the power and influence over the local context via CL teachers’ experience of AT. The findings may be situated within the context of Singapore CL teaching and assessment but would have wider resonance and application to other CHCs in general.

In summary, the above reviews showed that new assessment ideas from the ‘West’ had encountered much tension as they entered the CHCs. Teachers reported different degrees of struggle when they attempted to implement FA while still holding on to the traditional “examination-oriented” teaching. However, the use of CHCs as a lens to interpret the challenges and barriers faced by Asian teachers in assessment seems to have muted the “experiencers” voice and there may be a risk of stereotyping certain thinking and behaviour of certain group of people to CHCs when making conclusion, if their variation of such AT experience is not explored. The ‘better or poorer’ ways of experience AT are also not known from this perspective. Further, as Kennedy (2016) rightly pointed out, for the research to be more inclusive, other CHCs should be covered, and local teachers’ views should be continued to be expanded. Questions have been raised about the adequacy of treating culture as monolithic and deterministic in the CHC research agenda (Kennedy, 2016, p. 416-417). To this end, teachers’ experience of AT would have great implications and be valuable to the future research of CHCs.

2.9 Summary

From this review of literature on AT, I conclude that:

- (1) teachers experienced on-going AT due to the different assessment purposes they have to serve, which may conflict with one another. The notion of “balanced assessment” is not helpful in reconciling this tension;
- (2) the “two paradigms of assessment” caused AT because it is binary and thus put different assessment into a “zero-sum game” where “one’s gain is equal to the other’s loss”;
- (3) new assessment reforms or policies often intensified the AT experienced by teachers because they need to change their current assessment practices;
- (4) past studies on AT in Singapore have reported varied degrees of struggle by teachers when carrying out assessment due to multiple competing

considerations and that AT in actual situations is highly complex, especially when new assessment reforms are introduced. Different teachers respond to the AT in varied fashion and they are highly emotional about assessment change;

(5) given the historical and cultural background of the CHCs, the introduction of FA (or AfL) has met with massive challenges as their assessment cultures are repellent to change. However, teachers' voices regarding their experience of AT are largely not heard in the CHCs.

To sum up, despite numerous studies from the above five categories of literature on teachers' experience of tensions in assessment, the majority have not gone in-depth to explore how teachers experience and deal with these AT, from their viewpoints. In particular, previous studies on assessment tensions, stress, dilemmas, conflicts, contradictions and struggle in assessment have failed to recognise the importance of teachers' experience of AT and thus have not investigated this core educational problem from the teachers' perspective. Their collective and yet differing experiences, stories and voices of AT are largely not heard. Previous studies either identified the sources of AT or discussed about how teachers responded to these tensions. The distinctive ways in which how teachers experience AT in a 'better' or 'poorer' manner is still very much unknown.

In addition, compared to the amount of literature about teachers' experience of AT in western countries, there are far fewer studies in the pertinent research domain of CL teachers' experience of AT in a more eastern contexts, such as Singapore. Those that are conducted in the CHCs regarding AT only include Luk-Fong and Brennan (2010) and Yung (2001) in Hong Kong, Thanh Pham and Renshaw (2014) in Vietnam, Qi (2005) in China and Kim et al. (2013) in Singapore. Further, Rubdy and Tupas (2009), when reviewing research in applied linguistics and language teaching and learning in Singapore, reported that most studies done on language education limit themselves to describing the implementation of the proposed reforms and identifying the challenges met by teachers and students. Few of these researches investigated education policy and practice from the perspective of classroom teachers as central figures.

In fact, few have explored empirically how classroom teachers cope with and negotiate innovation that is top-down. In particular, Rubdy and Tupas (2009) also argued

that much of the local work on MTL instruction either exposit on the MOE's bilingual education initiatives or report in more detail on some of their relevant components through classroom strategies and innovative approaches to MTL teaching. The remaining work mainly focuses on how best to understand and implement recent state initiatives in bilingual education. To this end, the findings in this research add to the corpus of work on how new concepts in assessment from the 'West' (such as FA, AfL, alternative and sustainable assessment) are being taken on in Singapore, a CHC educational jurisdiction, by investigating teachers' experiences of AT.

This dissertation also takes the view that a subject-specific AT research will offer greater depth to understand how tension in assessment is experienced by the teachers in the CHCs. Particularly, it can significantly add to the literature of AT studies by offering these distinctive experiences in a single subject-specific context, alongside those of the general contexts that involved multiple subject disciplines. Future studies can then proceed to make a comparison with these experiences. By comparing these different experiences, we can determine which situations where we could offer teachers with 'better' experiences to manage AT are. We may also find out more about the limitations and potential of the contexts and conditions. When we compare (and contrast) subject-specific with general contexts AT studies systematically, we can discover their affinities and relationships which may ultimately make improvement in the ways which policy-makers and other stakeholders of education can offer teachers with better teaching and assessment conditions.

In sum, although past studies have identified AT as a core educational practice problem for teachers, open and candid revelations of how AT is experienced by teachers and effort devoted to such in-depth and rich investigation has not been given enough importance. This research may be the first of its kind in Singapore, as it builds a platform for CL teachers to describe their distinctive ways of experiencing AT. The collective differences in their experience of AT is captured and explored in greater depth. To this end, the methodology employed is qualitative, which does not only inform readers about what the AT are, but more importantly, it can capture the variation of ways of experiencing AT. It will be discussed in greater details in Chapter 3.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The first objective of this chapter is to show that the methodologies employed by past studies on AT, mainly quantitative approaches and case study, have not fully captured the in-depth and distinctive ways in which teachers experience and manage AT. The full picture of how teachers experienced AT in a specific context was largely not explored. Past studies have not investigated the collective differences in which teachers describe their real and close encounter with AT and their interrelatedness within. The employed methodology attempts to close this research gap. The second objective of this chapter is to explain the research method used that includes its basic concept, the investigation approach, data collection, analysis and interpretation. This chapter also hopes to convey an understanding of how a phenomenographic study in teachers' experience of AT is conducted, the rationale of employing this approach, the epistemological and ontological stands and what the implications are for the ways in which such study can be used to contribute and perhaps influence the research of AT.

To explore individuals' real experience of a phenomenon, I adopt an interpretivist paradigm which prefer to use qualitative research methods such as phenomenography that often uses interviews to investigate teachers' in-depth experience in educational research. According to Guba (1990, p.17), a paradigm or worldview is "a basic set of beliefs that guide action". Interpretivists propose an approach to researchers grounded in radically different assumptions and beliefs about the nature of the social world than those who propose positivism. In particular, interpretivists assert that the social world is fundamentally different from the natural world insofar as the social world is much more complex and full of different interpretable meanings. To this end, qualitative research is based on observations and investigations that can be recorded, discussed and interpreted but are not so easily counted or measured in numerical terms.

The philosophical assumptions of empirical research may be understood in terms of ontology and epistemology (Creswell, 2007). These research elements need to be clearly defined or the research is bound to suffer from validity issues. This is because

“the knowledge of these terms and their place in research is essential to understanding the research process as a whole” (Grix, 2002, p.175).

First, ontology refers to conceptions of the nature of reality. Blaikie (2000) defines ontology as “claims and assumptions that are made about the nature of social reality, claims about what exists, what it looks like, what units make it up and how these units interact with each other” (p.8). The ontology in this research assumes that the researcher and the participants view reality as subjective and thus there are many ways of explaining or interpreting a phenomenon. To this end, I am heavily dependent upon the use of face-to-face interviews with the participants to find out their authentic experience of AT. As teachers are dealing with daily assessment issues directly in schools, I believe that they will be able to give a more detailed, trustworthy and real account of the situation in contrast to other ways of collecting information such as large-scale surveys. Furthermore, during the interviews, both the participants and researcher have the opportunity to interact with one another closely. They are able to clarify any doubts pertaining to the topic discussed in greater depth. Because of the rich and in-depth interviews and discussions, the participants will probably be giving me multiple experience or evidence revolving around the research questions. In other words, there is a variation among their experiences which may be differentiated by how sophisticated (or ‘better’) their ways of experiencing AT. This perspective will give me a more extensive and collective picture of teachers’ experience of AT, which was lacking in previous studies.

Second, epistemology refers to how the researcher knows what he/she knows. According to Grix (2002, p.177), the word “epistemology” is derived from the Greek words *episteme* (knowledge) and *logos* (reason), epistemology focuses on the knowledge-gathering process. In this research, I am interested with questions such as “What is the knowledge derived from teachers’ experience of AT? How do I ascertain the validity of such knowledge about AT?” The epistemology in this research takes that reality needs to be interpreted and knowledge can be derived from teachers’ authentic experience. In phenomenography, an experience is a relationship between objects and subjects encompassing both the objective and subjective world. This is a “non-dualistic” ontological perspective (Marton, 2000).

From this perspective, individual teachers' accounts of their real experience of AT, as well as teachers' collective experiences, are valuable to our knowledge. In fact, teachers' qualitatively different ways of experiencing tension in assessment are insights which I wish to gain in this research.

3.2 Research methods commonly used in the studies of AT

This section reveals that the methodologies employed by past studies on AT, mainly quantitative approaches and case study, have not investigated the distinctive ways in which teachers experience and manage AT. As discussed in Chapter 2 (also see Appendix A), only a few of the studies on teachers' experiences of AT are quantitative and a handful of them used either the original or modified versions of Brown's questionnaire on Teachers' Conceptions of Assessment (TCoA). Some also used the original or modified versions of James and Pedder's (2006) model. However, qualitative research methods are still the most popular approach in past studies that reported tension in teachers' assessment practices. The data gathering and analysis in quantitative and qualitative research methods are fundamentally different. In qualitative research, the data collected are rich in description of people, places and conversations and not easily handled by statistical procedures. They are designed to explore topics in all their complexity in context (Bogdan & Biklen, 2007).

Denzin and Lincoln (2008) have also commented that qualitative researchers study phenomena in their natural settings, attempting to make sense of, or interpret the phenomenon in terms of the meanings people bring to them.

By investigating how teachers experience and respond to AT in natural settings with no intervention, readers can get a first-hand and in-depth understanding of how and why teachers react to AT in real-life. The findings may even allow readers to notice how teachers experience such special phenomenon that they might never have observed in an artificial setting. Indeed, qualitative approaches such as case studies (Bucci, 2002; Chang, 2006; Delanshere & Jones, 1999; Dixon & Haigh, 2009; Yung, 2001 among many others) have added depth and insights to the field of study. However, case studies are more interested in describing the group activities and not the shared patterns of behaviour in the group (Creswell, 2012). They usually focus on one setting, subject or event and do not focus on the relationship within and across the group.

Although “multicase studies” are in place, they are usually done to show generalizability or diversity issues and the cases are not done simultaneously (Bogdan & Biklen, 2007).

Further, there may be a methodological gap in some case studies when not much attention is paid to the diverse case selection strategy. This strategy is supposed to be able to explore two or more cases that exemplify diverse values. It is more likely to represent the full variation of the population. This coincides with my current reviews on methodology regarding the lack of use of diverse case selection strategy in AT research. This lack may imply that the inability to encompass a range of interrelated findings will probably be not able to enhance the representativeness of the sample of cases. In a way, phenomenography has great potential to contribute to the teachers’ experience of AT research in ways which quantitative methods and case study are not able to. This will be discussed further in this chapter.

The following section focuses on discussing the research methods employed by the past studies. It is argued that both quantitative and qualitative (particularly case study) approaches used by these past studies have not adequately explored the distinctive and yet interrelated ways in which teachers experience AT. To a large extent, these methods did not present and analyse satisfactorily the different aspects of a way of experiencing AT and how these aspects are interrelated. Teachers’ collective differences (or variation) in terms of the varying degree of sophistication (or ‘better’ or ‘poorer’ ways) of experiencing AT were also not sufficiently investigated.

3.2.1 The issues with quantitative research on AT

Past empirical studies have proven to be insightful of revealing the sources of AT teachers experienced and how they have been detrimental to teaching in both Singapore and elsewhere.

Out of the 62 empirical studies reviewed, only a few purely employed quantitative research methods that used Likert scale surveys or questionnaires. These studies usually involved bigger sample sizes and were conducted in the higher education or teacher education sector (see Appendix B). For example, Bonner and Chen (2009), used the Survey of Assessment Beliefs (SAB) to measure the 222 teacher candidates’ perceptions about grading practices in a United States urban university, found empirical evidence of AT between constructivist learning theory (FA) and principles of educational

measurement (SA). Factor analysis was conducted with the data collected by the six-point Likert scale questionnaires. Pearson correlation coefficients were used to assess the interrelationships among the measured variables and regression analysis was used to check the degree to which teachers' FA practice could be predicted by their characteristics and their beliefs about grading. Another example is the quantitative research conducted by Eren (2010) to examine 304 Turkish prospective teachers' values and practices in terms of their conceptions about assessment. Using the Teachers' Classroom Assessment Scale (TCAS) questionnaire, which comprised three factors (making learning explicit, promoting learning autonomy and performance orientation), the relationship between these teachers' assessment value and practice could be established and matched with Exploratory Factor Analyses (EFA). It was concluded that there were consonance and dissonance between these teachers' values and practices in assessment. In other words, AT were identified in these studies.

Two major issues related to methodology are identified from these quantitative studies. First, solely relying on Likert scale instruments such as questionnaires may not adequately assess or measure teachers' in-depth experience of assessment. More specifically, these closed-form instruments seem to assume that respondents view and interpret the items in a way similar to that of the surveys and questionnaires developers. For example, Brown's questionnaire on TCoA-III A has pre-determined conceptions for the teachers to choose. The four major conceptions of assessment are embedded in the questionnaires. The questionnaires were supposed to elicit teacher self-ratings for the four conceptions of assessment discussed in past literatures (i.e., Improvement, Student Accountability, School Accountability and Irrelevant) by Brown (2003, 2004, 2008). Teachers' individual voices and stories are not heard directly from their own experiences. James and Pedder's (2006) original and modified versions of the classroom assessment and learning questionnaire also have the issue that they could not find out teachers' in-depth experiences of assessment. Further, teachers' actual experience of assessment may not be fully represented and elucidated based on numerical values of one to five (or six). In other words, the complexity and richness of teachers' experiences of AT are not studied in-depth.

The second issue concerns the language used in the questionnaires. There is an assumption that the potential respondents would fully comprehend and are

comfortable with reading and writing in the language of the questionnaires at the same level (Vogt, 2007). Unlike face-to-face survey or interview, there is no room for the respondents to make in-depth explanations and offer clarifications. These in-depth investigations are necessary in order for me to find out the nuances and the distinctively different ways in which CL teachers experience AT.

In educational research, the planning, conducting, analysis and evaluation of quantitative and qualitative studies are fundamentally different. For instance, while quantitative research uses more closed-ended approach (such as Likert's scales), qualitative research prefers open-ended approach (such as face-to-face interviews) (Creswell, 2012). Indeed, researching into teachers' experiences of assessment will be problematic if only quantitative research methods are employed because I am interested to find out the complexity in teachers' minds relating to their experience in assessment. These qualitative studies usually attempt to develop a detailed understanding of a central phenomenon (in this case AT) and would require a smaller number of participants' in-depth views. The credibility and accuracy of the research will be ensured if qualitative research methods are used to investigate teachers' experience of AT in this research.

3.2.2 The issues with case study

Case study is one of the main approaches used by qualitative researchers to explore teachers' experiences of assessment. Out of the 62 reviewed studies that have employed qualitative research methods, 13 have explicitly labelled themselves as "case study". However, based on the definitions and discussions below, it was found that majority of the rest of the 49 studies could also be categorised as "case studies". This is because the scale, approach, sample selection and research tools in the latter studies resemble those in the former 13 studies.

So far, there is no consensus on the definitions and characteristics of a case study (Blatter, 2008; Thomas, 2011; Yin, 2009). Because of this, different kinds of case studies were enumerated by different scholars depending on their epistemological starting points (Bassegy, 1999; de Vaus, 2001; George & Bennett, 2005; Merriam, 1988; Stenhouse, 1985; Stake, 1995; Thomas, 2011; Yin, 1994, 2009). Although the 49 studies did not claim to have employed the case study, they exhibited various attributes of this

research approach based on past definitions. In educational setting, case studies can be understood as a critical enquiry aimed at informing educational judgements and decisions in order to improve educational action (Bassey, 1999, p.39). They are usually conducted by educationists and have the stated intention to inform practitioners and other stake-holders of education (Bassey, 1999).

One of the advantages for using case studies is their uniqueness and capacity for developing an in-depth understanding of complexity in particular contexts. A corresponding disadvantage often encountered by researchers is “the difficulty of generalising from a single case” (Simons, 1996, p. 225). Indeed, from the 40 reviewed studies, uniqueness seems to be the most prominent characteristic of each case study. To this end, generalisation in case study is paradoxical (Bassey, 1999; Simons, 1996) and thus debatable (Tripp, 1985). On one hand, one of the main objectives of conducting a case study is to establish generalisations about the wider population to which that unit belongs in a phenomenon (Cohen & Manion, 1989; MacDonald & Walker 1975; Seawright & Gerring, 2008; Tripp, 1985). On the other hand, many have criticised a single study that it has little basis for scientific generalisation (Mookherji & LaFond, 2013; White & Phillips, 2012; Yin, 2009). Seawright and Gerring (2008) opined that case selection is the primary task for the case study researcher as the data analysis that follows will be largely affected by the selection too. They then went on to recommend seven case selection procedures or types that include typical, diverse, extreme, deviant, influential, most similar, and most different cases. These seven types differ in their use and representativeness.

A more in-depth “re-review” on the 62 qualitative studies revealed that none had employed the “diverse case selection strategy”. For example, the studies done by Arkoudis & O’Loughlin (2004), Hammerness (2004), Thanh Pham and Renshaw (2014), Cella (2005), Maloley (2008), Arden (2010), Tan and Leong (2014) and Yin (2013) have only involved one or two participant(s). This means that they are unlikely to be representative in the minimal sense of representing the full population. At this point, the “diverse case selection strategy” deserves our attention. According to Seawright and Gerring (2008), the main objective of the diverse case selection strategy is to achieve maximum variance along relevant dimensions. They further explained that this method had not received much attention on the part of qualitative methodologists, hence the

absence of a generally recognized name (Seawright & Gerring, 2008, p. 307). Their observation surfaces a possible gap in many “case study” researches because not enough attention is paid to the diverse case selection strategy. This coincides with my above re-review on methodology regarding the lack of use of diverse case selection strategy in past AT studies. The reviewed case studies that are similar to the context of this research include the research by Davison (2004), Luk-Fong and Brennan (2010) and Thanh-Pham and Renshaw (2014). These are done in the CHCs where the educational culture and heritage resemble the case of Singapore. Another group focus on teachers’ experience of AT in a subject-specific context, particularly on teachers’ experience of language assessment, which is similar to my subject-specific context of exploring CL teachers’ experience of AT. They include Arkoudis and O’Loughlin (2004), Chang (2006), Deng and Carless (2010), Lumley (2002) McKay and Brindley (2007), Norashikin et al., (2010) Shihiba (2011), Taylor and Wallace (1990). These case studies have not employed the diverse case selection strategy in their investigations.

The diverse cases are likely to be representative in the minimal sense of representing the full variation of the population. When there is an inability to encompass a wider range of differences (or variation) in case studies, it may go on to affect the representativeness of the sample of cases. When this happens, there may be a risk of stereotyping certain behaviours to some biased frame of explanation. For example, in my practice context, there may be a tendency to use the CHCs or Asian values to analyse and explain CL teachers’ thinking and actions due to the lack of variation. To this end, the structure of ‘better’ and ‘poorer’ ways of experiencing a phenomenon in phenomenography will be very useful because there is a strong perception among CL teachers that AT is negative. There is also a tendency in my subject discipline to perceive CL teachers to be experiencing much tension because they seemed to be very “exam-oriented” in their pedagogy and assessment practice, especially when new assessment ideas from the ‘West’ are being introduced. Such cases are in danger of standardising how CL teachers in my practice context would experience and manage AT, which may marginalise certain groups of teachers in practice. For this reason, the selected samples in this phenomenographic research need to be diverse and varied so that such causal generalisations could be avoided. Further, when the distinctive differences of how teachers experience AT is captured, this will have a direct impact on

the in-depth findings of how some of these AT experiences may be considered as 'better' or 'poorer' from teachers' perspectives. This is where the research approach, phenomenography, can significantly close this methodological gap in the study of teachers' experience of AT. Phenomenography can capture and explore teachers' collective experience of AT in terms of their distinctive differences (or variations) and their interrelatedness by purposefully selecting teachers who are different in their profiles. This approach will be discussed in greater detail.

3.3 Unique contributions of phenomenography to explore teachers' experience of AT

Phenomenography has been widely used for investigating teachers' experiences and conceptions of teaching, learning as well as assessment in educational studies (Akerlind, 2005; 2008; Bowden, 1996; Denee & Brown, 2010; Marton, 1981, 1986; Marton & Booth, 1996; Marton & Pong, 2005; Tan & Prosser, 2004; Tan, 2010, 2012, 2013; Wang, Kao & Lin, 2010). Spearheaded by Marton in the 1970s at the University of Gothenburg in Sweden, phenomenography as an internationally valued research approach was developed. It was then formally introduced in 1981 by Marton in the educational field. Marton (1995, p.166) defines phenomenography as a research approach that "aims to reveal the qualitatively different ways in which something is experienced".

Hasselgren and Beach (1997) have highly commended phenomenography and concluded that Gothenburg phenomenography is a very productive research approach which can be developed in five different ways. According to Johansson, Marton and Svensson (1985), a conception in phenomenography is a particular manner in which people experience and understand some phenomenon. A person's conception of something is actually a relationship built between the internal self and the external world. As such, everyone's conceptions of something may not be the same.

In this section, it is argued that phenomenography offers a few contributions toward teachers' experience of AT research. Firstly, phenomenography allows the practitioners to describe their real experiences of AT. It can help participants to describe the distinctive character of their AT experiences as they appeared to them and present what is most important in that sense. This approach seeks to capture and investigate

how the phenomenon in the world appears to other people and how these experiences are related to each other.

Secondly, the phenomenographic research will be significant because to date, little information has been available regarding CL teachers' varied and yet collective experiences of AT in terms of their differences and their interrelated relationships within. It is only when people recognise and are made aware of the distinctively different ways of experiencing AT may then start to appreciate how the different approaches to understand AT in relation to each other. In fact, by encompassing and understanding the full progression (or hierarchy) of the different ways of experiencing AT, it is likely to enhance the richness and representativeness of the sample of cases chosen, the stakeholders of education may then widen the foundation of common comprehension and meaning of this tension. They will be informed of the 'better' (and 'poorer') ways of experiencing AT from teachers' perspective.

Thirdly, as individual's experience of a phenomenon is always embedded in a context, I argue that there is a need to pay attention to human beings' "structure of awareness" of experience. When an individual experiences a phenomenon, certain things will be at the fore and others will be backgrounded. In Marton and Booth's (1997) words, "there are different degrees of how figural, thematised, or explicit things or aspects are in our awareness" (p.98). In contrast to the binary perspective of the "two contrasting paradigms of assessment" that is always seen as a "zero-sum game" in assessment practice, the "structure of awareness" of AT in phenomenography can provide readers with an alternative and stable lens of explanation on how AT may be experienced, as well as how these experiences are related. Through this structure, the complexity of AT may be better understood since the binary perspective of the "two contrasting paradigms of assessment" has been limited to fully explain or depict teachers' lived experience of AT.

This research is particularly useful in my practice context where AT has often deemed to be negative. This "structure" of how teachers experience AT in past case studies is not heard of and it can be used to develop a deeper understanding of teachers' experience of AT beyond negativity. The "structure of awareness" of AT provides readers with a wider and deeper understanding of the distinctiveness of each way of experiencing AT, as well as the relationships between these different experiences. Case

studies did not seem to be able to offer such a possibility. More details on the “structure of awareness” will be discussed later.

Phenomenographic researchers strongly believe that improvement of complex activities such as teachers’ assessment practice requires a rich and in-depth understanding of the interpretive nature of the relationship between interpretations of the experienced phenomenon and their actions. Subsequently, with this insight, future assessment policies planning, initiatives, reviews or reforms may be built based on these grounded and experienced contexts of teachers’ experience and conceptions. This research may also set the stage for a series of on-going recommendations, adjustments and even interventions around the redesigning of language curriculum, teaching and assessment system and policy in Singapore.

3.3.1 Basic concepts of phenomenography

Marton (1994) defines phenomenography as: “the empirical study of the limited number of qualitatively different ways in which various phenomena in, and aspects of, the world around us are experienced, conceptualised, understood, perceived and apprehended” (p.4425). Marton and Booth (1997) unpacked that, “a way of experiencing something” is a way of discerning something from, and relating it to, a context. The meaning of something for someone at a particular point in time corresponds to the pattern of parts or aspects that are discerned and are simultaneously objects of focal awareness.

Marton (1995, p.166) further explains that phenomenography “aims to reveal the qualitatively different ways in which something is experienced”. So the interaction between the individual and the phenomenon is the core here because one cannot function without the other. From an ontological perspective, it is an “interpretivistic view” which sees reality as “non-dualistic”. Marton (2000) explained that in the existing world, different individuals experiencing a particular phenomenon will reveal their varying experience. The descriptions of their experience are based on ways of understanding the particular phenomenon. However, “each phenomenon, concept or principle can be understood in a limited number of qualitatively different ways” (Marton, 1986, p.30). In this research, the purposefully selected CL teachers from varying

backgrounds and teacher trainings will describe their assessment experience. It assumes that a limited number of variation of conceptions (in a related form) can be identified.

3.3.2 Unique features of phenomenography

The unique feature of phenomenography is that it enables readers from different contexts to appreciate how a phenomenon may be experienced in different manners (Tan 2008a).

Tan (2010), when arguing why phenomenography is suitable for investigating the meanings of alternative assessment experienced by Singapore teachers, highlighted that there are theoretical, analytical and pedagogical benefits for utilising this research approach. In the same way as Tan (2009, 2010) has argued, I argue that phenomenography (or variation theory) will be beneficial to readers for understanding teachers' experience of AT brought about by the system and educational reforms. Phenomenography adds to the theoretical, analytical and pedagogical dimensions of understanding the tensions in assessment brought about by national high-stakes examination system and educational reforms or initiatives in Singapore from the perspective of those teachers who have experienced it. Such dimensions depict how the experience of assessment and education policies are theorised by individuals (theoretical aspect), analysed to present the different aspects of a way of experiencing AT and how these aspects are related (analytical aspect). The distinctive research benefit of knowing teachers' ways of experiencing AT is that we could suggest how better ways of understanding and initiating assessment reforms or changes may be achieved (pedagogical aspect) so that 'better' ways of experiencing AT can be brought about.

Indeed, as Marton and Booth (1997) have previously argued:

in order to make sense of how people handle problems, situations, the world, we have to understand the way in which they experience the problems, the situations, the world, that they are handling or in relation to which they are acting. Accordingly, a capability for acting in a certain way reflects a capability experiencing something in a certain way. The latter does not cause the former, but they are

logically intertwined. You cannot act other than in relation to the world as you experience it (p.111).

From above, I can better understand Marton's (1981) earlier work where he argued in detail why he preferred to explore the "second-order perspective". He explained that the "first-order perspective" describes the world without taking into account individual's way of experiencing it. On the other hand, the "second-order perspective" is focused on people's ways of experiencing, interpreting, understanding, apprehending, perceiving or conceptualising "various aspects of reality" (Marton, 1981, p.178). This view is particularly useful for this research on teachers' experiences of assessment because it is consistent with the research question which aims to explore how CL teachers describe their experience of AT in the context of Singapore.

3.3.3 Aims of research and sample variation

The main aim of this research is to provide a conceptual description and analysis of the various views CL teachers in Singapore have of the actual AT they encountered in practice. Phenomenography enables me to investigate the various types of AT experienced by the teachers. In order to maximize the possibility of obtaining distinctive variation, samples were carefully selected.

Assume there are two CL teachers teaching in two similar profile secondary schools in Singapore. Imagine that they have similar educational backgrounds, experienced the teacher trainings, practicum, teaching and assessment in exactly the same way. They understand their education system and know the problems and tensions brought by the assessment policy and culture. They encounter the assessment situations in exactly the same way and they feel that it was also equally important for both to face the problems and challenges embedded. However, they also know the limitations they have as a classroom teacher in terms of assessment practices. In other words, their experiences of the assessment problems, challenges, tensions and the situations were identical. Could we imagine that one of them will try his or her best to pro-actively promote formative classroom assessment while the other teacher totally ignores the benefits brought by assessment in language learning? This is very unlikely because the two teachers who have had experienced assessment in a similar way,

cannot approach assessment issues or situations in such a diverse way. This is where phenomenographic methodology will fit in seamlessly in describing and exploring CL teachers' experience of AT in Singapore because of their varied backgrounds. The complexity and richness of the AT experienced will be well captured. The total population of CL teachers in Singapore is about 3500, with 1200 in secondary schools (Singapore Centre for Chinese Language, 2010). CL teachers are a very special group in the teaching fraternity of Singapore because they came from very different academic backgrounds and received different teaching trainings. For example, there are very senior teachers who graduated from the Nanyang University (a Chinese medium university) in the 1960s-70s and have taught CL for over thirty years at L1 level. There are beginning teachers who just graduated from National University of Singapore (NUS), Nanyang Technological University (NTU) or National Institute of Education (NIE). There are also foreign CL teachers from China, Taiwan and Malaysia teaching in Singapore.

Marton and Booth (1997) advocated maximizing the "conceptual variations" of participants to ensure sufficient and rich data so that an optimal set of categories can be derived. As CL teachers are very diverse in their backgrounds, it is argued that phenomenography will provide much room for the researcher to explore the "variations" of how they experienced assessment. In this research, CL teachers' characteristics (such as their academic backgrounds, teacher training and years of service) have been documented so that the spread of characteristics is intended to maximize the "conceptual variations" in the data. See Table 3.4.1 below for details. The main aim of a phenomenographic study is to investigate the variation in the ways a phenomenon is experienced. As such, if the researcher is careless in considering and selecting the appropriate samples for investigation, many critiques could be heard from the research community. In such a case, there could be a possibility that the full extent of the variation of how teachers experience AT may not be assured.

In this research, I intentionally selected the participants to explore and understand teachers' experiences of assessment. I purposefully selected CL teachers who have a large amount of variation in their profiles. Although Newton and Newton (2009) believe that phenomenographical investigations will require "interviews with between 12 and 20 people" (p. 50), this research has selected just nine CL teachers for interviews. A handful of the previous phenomenographic studies has also involved

samples which are fewer than 12. For example, research done by Tan and Prosser (2004) only conducted in-depth interviews with seven academics regarding the different conceptions of grade descriptors. Another phenomenographic study done by Deneen and Brown (2011) to investigate teachers' conception on the nature and purpose of assessment only used six participants from a population of both in-service and pre-service teachers enrolled in New York City educational course. They justified that although small in number, the sample would already provide representation of a diverse population of practicing and pre-service teachers, male and female, naturalized and African Americans, as well as teachers from Eastern Europe and Russia.

Although limited in number, the samples in this research have provided a comprehensive representation of a variation of CL teachers from greenhorn to experienced, from local to foreign. In addition, the overall ability to do an in-depth qualitative research diminishes if the sample size and sites get larger (Creswell, 2012).

3.3.4 The structure of awareness and outcome space

Different dimensions of conceptions regarding AT may emerge from the interview data. For example, this research included some open-ended questions that will open up teachers' conversations on the considerations they have when making assessment decisions. Teachers may give their opinions and contribute their versions of thinking and eventually will enable this research to construct not just a set of different meanings toward AT, but also a logically inclusive structure that pulls all these meanings together in a hierarchical manner. As a whole, the different dimensions of the AT experienced by the teachers could emerge from the data, which will eventually form the outcome space. This outcome space will contain a structure where all the teachers' interviews can be roughly positioned. These are the focuses of the awareness of teachers' experience of AT. The structure of awareness denotes the relationship "between" and "among" different categories of description. As shown in Table 3.3.4 below, the structure of awareness has three main components to it and each either belongs to the internal or external horizon. The theme is the focal of awareness in the structure. It is a person's general awareness of the phenomenon experienced. Figure 3.3.4 illustrates the relationships among different components of the structure of awareness.

Components	Definition	Horizon
Focal awareness (theme)	The phenomenon experienced	Internal
Thematic field	Those aspects of the experienced world that are related to the object and in which it is embedded	External
Margin	All those aspects which coexist with the theme without being related to it	External

Table 3.3.4 Structure of awareness (Bowden & Marton, 1998; Marton, 2000; Marton & Booth, 1997; Tan, 2010)

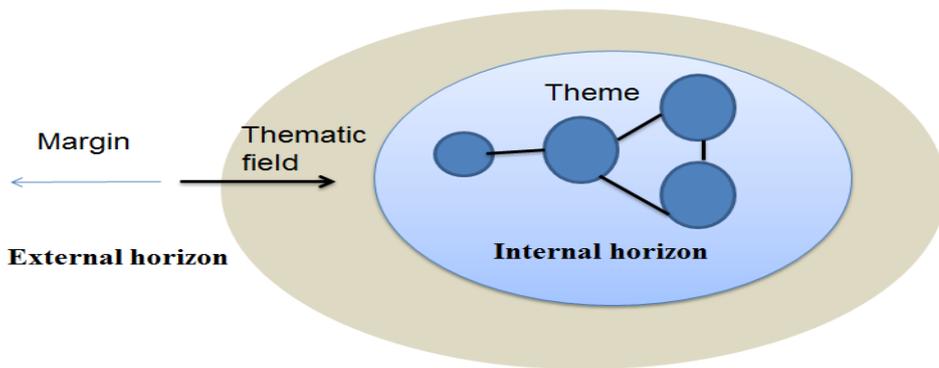


Figure 3.3.4 Structure of awareness

This structure is in line with what Marton and Booth (1997) posited when they talked about the assumption of phenomenography. They believed that the qualitatively different ways of experiencing a phenomenon would be logically related to one another, with increasing complexity where the qualitatively different ways of experiencing the phenomenon would form as “subsets” of the component parts. This structure of awareness offers a new way of explaining how CL teachers, in the contexts they are in, may relate to the same phenomenon of experiencing AT. This model will not predetermine CL teachers’ experience of AT to be within the scope of issues reviewed in Chapter 1 and 2 which included conception of assessment, core educational problems associated with AT and so on. Instead, the critical aspects of how teachers experience AT can be discerned from this phenomenographical research.

The unique contribution of this research, in the light of all the quantitative approaches and some qualitative approaches such as case study, can be seen from the “outcome space”. This research aims to explore and construct not just a set of different meanings made by the participants, but a logically inclusive structure relating these different meanings. In phenomenography, the qualitatively different ways of experiencing the AT will be constructed into “categories of description” and are seen as representing a structured set, which is termed as the “outcome space”. Even though the same phenomenon may be experienced and conceptualized differently by different people and under various contexts, phenomenography can provide a lens looking into collective human experience of a phenomenon in a holistic way. This is because the “outcome space” essentially represents the whole range of possible but yet limited ways of experiencing the phenomenon, at this particular point in time, from the selected group of people in a collective manner. This is where the complexity of AT could be revealed in a structure which previous studies have not to date adequately addressed.

Typically, the qualitatively different ways of experiencing the AT may be described as “conceptions” and they are represented by the “categories of description”. According to Marton and Booth (1997), there are three criteria that need to be met in order to be qualified as one set of categories of description. Firstly, each category of description should signify a distinctly different aspect of the experienced phenomenon. Secondly, different categories of description should form a logical relationship in a hierarchical fashion. Thirdly, a minimum number of different categories must be formed and described in order to differentiate the variation of experience in the outcome space. Essentially, this research seeks to explore CL teachers’ ways of experiencing AT such that: (1) Each category of description denotes a qualitatively different way of experiencing AT by the CL teachers. The description of their similarities and differences would divide these categories distinctively. (2) Each category of description represents a progression of awareness of the critical aspects of experiencing AT and eventually form a logical relationship together. (3) The outcome space is limited to a few ways of experiencing AT by the CL teachers. This number of categories describes the critical aspects of different ways of experiencing AT from the nine participants.

3.4 Data Sources and Collection

While the above sections dealt mostly with the theoretical underpinnings of phenomenography, the following sections discuss this methodology in practice, particularly to describe the actual steps drawn from those theoretical premises. In this research, data was collected from in-depth personal interviews.

3.4.1 Identifying participants

In Singapore, CL teachers vary in their nationality, academic and teacher-training backgrounds (see table below for the profiles of the participants).

<i>S/N</i>	<i>Teacher</i>	<i>Age</i>	<i>Gender</i>	<i>Background</i>	<i>Years of teaching</i>	<i>Teacher training</i>	<i>Assessment training (pre-service)</i>
1	A	40	F	China	12 years	China (扬州大学)	No
2	B	35	M	Malaysia	10 years	Local (NIE)	Yes
3	C	43	F	Local Singaporean	22 years	Local (NIE)	No
4	D	30	F	China	5 years	China (厦门大学)	No
5	E	42	F	Local Singaporean	14 years	Local (NIE)	Yes
6	F	34	F	Taiwan	9 years	Taiwan (清华大学)	Yes
7	G	29	M	Malaysia	4 years	Local (NIE)	Yes
8	H	28	F	Local Singaporean	4 years	Local (NIE)	Yes
9	I	60+	F	Local Singaporean	40 years	Local (Nanda)	No

Table 3.4.1: Teachers' Basic Profile (as of 2014)

From Table 3.4.1, CL teachers' actual names are not disclosed. Their consents to participate in this research were sought. Except for Teacher B and H, the rest are teaching in secondary schools of different profiles. The table also shows if they have received pre-service assessment training. This difference in background naturally

provided an opportunity for me to select the participants from a broad range of experience of tension in assessment in search of variation. In addition, by approaching CL teachers with varied years of service from a broad spectrum of work places in Singapore that included both “neighbourhood” and “elite” schools (the former usually refers to government-aided schools that are not so famous in terms of students’ academic performance; the latter usually refers to autonomous or independent schools that are more famous for their academic performance), as well as education officers from the MOE headquarters, I sought to ensure there was a diverse range of practices and thus experiences represented in the research. It is to note that education officers from the MOE headquarters are also previously from schools. They joined the headquarters to learn how to plan and develop the centralised CL curriculum. They would usually return to schools after two to three years in MOE headquarters.

However, there is a need for me to delimit the precise experience and investigate those that are relevant to my thesis definition of AT. I have to ensure that what the participants experienced and shared with me is relevant to my research. This was carefully dealt with during the interviews by not drifting into teaching practices and experiences which were not within the scope of the definition of AT. I had defined AT as “teachers’ feelings of struggle with competing considerations in their assessment practices”. The critical element in this definition is that CL teachers must express their assessment experience with the feeling of “tension”, “dilemmas”, “conflicts” or “struggle” in practice because of certain considerations when making assessment decision. There were several experiences which have been often described by teachers as AT but are excluded in this research. The first one is that of teachers’ account of other teachers’ experiences of AT instead of theirs. The second experience I excluded is their national examination marking duties. These experiences are supposed to be confidential and not be disclosed to any other party.

CL teachers from three secondary schools of different profiles and the Ministry of Education (MOE) headquarters were selected as participants. They were all approached by email and telephone. While the first email sought their preliminary permission to carry out the interviews, the second email explained the details of my research. Documents attached in the second email included the basic information about my research for participant, consent form, interview protocol form, as well as the

Institutional Review Board (IRB) approval letter from the MOE. These documents were used during the interviews. A sample of the emails and attached documents is set out in Appendix C and D respectively. While some of the participants were my friends in the teaching profession, others were recommended by my colleagues.

3.4.2 Planning and conducting the interviews

This research explores the phenomenon of teachers' tension in assessment through their collective experiences. These experiences are understood from the interviews and were then analysed for the iterative and logical meanings.

A semi-structured interview is the primary data collection method in phenomenography (Marton, 1986) and it is "non-directive" (Walsh, 2000). In fact, Booth (1992) explains that by using a few pre-determined questions, the purpose of semi-structured interviews is actually to probe into the participants' understanding of certain phenomenon from different angles so that a full exploration of understanding is made possible. This is very evident when I reviewed the literatures in phenomenographic research (such as Deneen & Brown, 2011; Wang, Kao & Lin, 2010) that explored teachers' experiences and conceptions of assessment. All the interviews were semi-structured and there was sufficient room for all participants to express their experience but at the same time be focused on a specific topic. Tan's (2010, 2012, 2013) phenomenographic studies on teachers' experiences of alternative assessment in Singapore also placed a heavy emphasis on semi-structured interviews. He explained in detail the rationale of using semi-structured interviews and the structure of the interviews. The studies follow closely to what Bowden (2000) has suggested in phenomenographic interview:

The phenomenographic interview has a focus—the way in which interviewees understand the chosen concept—and this focus is maintained throughout the interview. Interviewees are encouraged to express their qualitative understanding of the phenomenon under investigation. The researcher may ask interviewees to clarify what they have said, and ask them to explain their meaning further (p.9-10)

Similarly, in this research, there are over-arching questions and pre-determined questions (see interview protocol in Appendix E for details) which were used when the participants did not touch on the crucial aspects of the phenomenon. These reviewed studies, together with many other similar phenomenographic studies, lend support for the employment of interviews in this research to explore CL teachers' experience of AT. Essentially, the "seven stages of an interview investigation" (thematizing, designing, interviewing, transcribing, analyzing, verifying and reporting) proposed by Kvale (1996) also provided a systematic approach about how a high standard phenomenographic interview can be conducted.

In the same vein, the interview approach for this research is similar to what the above-mentioned researchers did or advocated. The interview protocol was designed to take care of the instructions for the process of the interviews and the questions to be used, as well as how the data should be captured. Previous phenomenographic studies reviewed in this chapter (Akerlind, 2005; 2008; Bowden, 1996; 200; Denee & Brown, 2010; Marton, 1981, 1986; Marton & Booth, 1996; Marton & Pong, 2005; Prosser 2000; Tan & Prosser, 2004; Tan, 2010, 2012, 2013; Trigwell, 2000; Wang, Kao & Lin, 2010) have provided good samples for the interview protocol.

The phenomenographic interviews have a focus in the sense that it was underpinned by the research question. The participants were led by me with pre-designed overall interview questions and sub-questions to reiterate their experience of AT. To effectively capture the experience of the teachers, I adhered to the following:

- (1) The questions were asked in Mandarin Chinese so that the CL teachers would not have any language barriers to fully comprehend the research, as well as to clearly express their views. It is also to minimise misunderstanding, reliability and validity issues brought about by language factors (Geisinger, 1994). More importantly, this research highly values their original and authentic voices and experience. This can only be transmitted by their most comfortable and natural language.
- (2) The data was first recorded in CL and subsequently translated into English when reporting so that the essence of the teachers' experience would be captured at the first opportunity.

(3) The questions were open-ended so that the researcher would not be dictating the aspects of the question which were most relevant to the teachers. They were free to offer any possible response to the questions.

(4) The questions were designed to be diagnostic so that they can better reveal the different ways of understanding the phenomenon.

However, adherence to the above posed certain challenges. For example, the translation between English and CL done by me may not be as accurate as the professional translator. Furthermore, certain descriptions and meanings given by the teachers might be lost in translation. Indeed, it is always challenging to achieve conceptual equivalence between two languages in cross-language research for there is always translator bias (Choi, Kushner, Mill & Lai, 2012). As this research was conducted in a language that is not the researcher's reporting language, therefore, translation required additional knowledge and work. Nevertheless, the main translation was done by me as I was more familiar with the context, culture and experience of CL education in Singapore. Further, by consulting and communicating with more qualified translators and to practise both "forward translation" as well as "backward translations" (Choi, Kushner, Mill & Lai, 2012, p.661) have helped to reduce potential threats to the validity of the data and interpretation.

As many phenomenographic researchers have noted, the effectiveness of an interview depends very much on the interviewer's ability to design and ask questions that may increase the chances of a full exploration of the interviewees' understanding of the phenomenon. In this research, I designed a variety of semi-structured interview questions and employed multiple ways of asking these questions if the need arose. According to Marguerite, Spaulding and Voegtle (2010), semi-structured interviews are typically planned prior to the actual interview. The researcher usually can change the order (structure) of the planned questions, omit questions and change some of the wording of the questions based on what actually happens during the interview. This is to probe deeper into the unexpected issues that emerge.

As defined earlier, AT refers to "teachers' feelings of struggle with competing considerations in their assessment practices", therefore I designed five types of questions to guide participants to become aware of their experiences with AT in practice,

as well as to diagnose and reveal their different ways of experiencing this tension. A well-designed interview would allow the participants to reveal something significant about themselves which would be in greater depth. Because of the questions asked, they would sometimes suddenly realise their own thinking and feelings toward certain phenomenon (Bowden, 2005). How could I achieve this objective and at the same time ensure that the ethics requirements and the comfort level, as well as their willingness to cooperate, were not compromised? The types of question asked would be crucial. There were basically three main components to the phenomenon and these questions were supposed to keep the participants focused on describing their experience of AT. First, this experience must be from the participants' assessment practices. For this point, participants might be asked to bring along their assessment artifacts for discussion which would add richness and depth to the discussions. Second, there were multiple considerations when deciding what to practise in assessment. Third, there was a feeling of struggle at some point in time when planning or conducting assessment.

In summary, each of the following five types of interview questions has a specific aim of understanding participants' experience of the phenomenon. Table 3.4.2 illustrates with examples how I selected and employed the five different types of questions during the interviews. These five types of question are crafted based on the advice given by Bowden (2000, p.9) on how a phenomenographic interview should be carried out in terms of the different kinds of conversation involved.

Type	Main aim	Examples of interview question
1	Opening up the conversation and discussion (warm up questions)	<ul style="list-style-type: none"> • What is your assessment practice and how do you carry out assessment? • How often do you carry out assessment? • What kind of assessment do you use in your classrooms and why?
2	Underpin the focal awareness of the participant	<ul style="list-style-type: none"> • In your assessment practice, what are your considerations: such as time, resources, accountability, students' needs etc?

3	A small selection of prepared questions to begin another new inquiry	<ul style="list-style-type: none"> • How do you feel when carrying out these assessments? • How do you deal with these different considerations in assessment? • If you were given the freedom to design assessment, how would it look like? Will it be different from your current practice? What will be the differences? Why? • Do you think there are better assessment practices?
4	A small selection of prepared sub-questions to further probe, confirm, interrogate or for clarifications	<ul style="list-style-type: none"> • Could you explain/elaborate that further? • Could you give me some examples? • What do you mean by that? • By saying……do you actually mean/imply that……?
5	Concluding and ending the interview	<ul style="list-style-type: none"> • That is about all, do you have any other things to add or comment?

Table 3.4.2: Types of interview questions used in a single interview

The very first type (Type 1) of interview questions was to open up a conversation with the participants but would still stay within the context of the research question. As such, these questions were directed at more general assessment issues such as “What is your assessment practice and how do you carry out assessment? How often do you carry out assessment?” These questions directed participants’ attention and allowed them to focus on their assessment experiences which were highly relevant to the delimited phenomenon and context. The research question of this research is “What are the qualitatively different ways in which teachers experience AT?” And the definition of AT is “teachers’ feelings of struggle with competing considerations in their assessment practices”. Therefore, Type 2 is the primary question that underpins the participants’ focal awareness, helping them to stay in focus on their experienced phenomenon. The

question “In your assessment practice, what are your considerations (such as time, resources, accountability, students’ needs etc)?” had helped the participants stay relevant to the research question. Besides these “warm-up” and “primary” questions, Type 3 consisted of a small selection of prepared questions to begin another new inquiry, which were intended to cover the scope of CL teachers’ awareness of AT. These questions encompassed the scope or aspects of AT which I wished the participants could describe their experience and thus their conceptualisations of it. These questions were particularly useful when the participants did not already talk about them during the interviews. Essentially, Type 3 questions were used to serve the following three functions:

(1) Exploring teachers’ feelings in their assessment practice: This was one of the most important aspects of the interviews which required CL teachers to describe their feelings or emotions as they experienced assessment. I prepared questions to directly probe what the teachers discerned to be important emotionally. The question I designed was straightforward: “How do you feel when carrying out these assessments?” There were a variety of responses from this question. While some were positive (such as feeling accomplished or happy), others shown more negative experiences (such as feeling stressful, a waste of time, being ‘pulled’). There were also some neutral responses such as “no feeling”, “a mixed of feelings” or “I didn’t bother about feelings”. To a large extent, these reported feelings demonstrated the complexity of teachers’ experience with assessment and confirmed they were experiencing different degrees of tension in assessment.

(2) Setting up the exact context of experiencing AT: I wanted to find out the qualitatively different ways in which CL teachers experience tension in assessment and so the conversation needed to be situated in their actual assessment practice. Accordingly, these participants had to recount their experience which must be important and meaningful to them. By doing so, I am establishing the context in which CL teachers experienced AT. As important as how I delimit the context for this research, participants’ current context for the phenomenon was also vital for analysing the transcripts. The question I asked was: “How do you deal with these different considerations in

assessment?” This question was effective in getting the teachers to articulate their practical concerns.

(3) Inviting teachers to reflect on their assessment practice: In the interview, I aimed to make participants reflect deeply on what tension in assessment really meant to them at the point it was experienced. As such, I designed the following few questions:

If you were given the freedom to design assessment, how would it look like? Will it be different from your current practice? What will be the differences? Why? Do you think there are better assessment practices?

I observed that these questions often made participants pause for a while, or they requested me to repeat the question before they responded. This showed that they needed time to think deeply into the questions and gave me responses which were deemed to be important.

In order to help the participants get into the right condition to describe and reflect on their experience of AT, they were often asked Type 4 questions to clarify, further explain or elaborate their statements. Under each of different type of question, there were also a number of similar and reiterative questions. For example, I had questioned them further with “Could you explain that further?”, “What do you mean by that?” These questions aimed to get the participants to think deeper into their statements and descriptions of their experiences which would enable them to explain their understanding more fully (Bowden, 2000). Occasionally, I also had to make a more detailed inquiry into the responses provided and probe further by rephrasing the words or sentences given by the participants. The question was “By saying.....do you actually mean/imply that.....?” This question aimed to allow participants to express variation in their experience of AT. In other words, I was creating another opportunity to allow the participants to describe their experience of the phenomenon in their own words. By doing so, I would not impose any description for the participants (Entwistle, 1997; Bowden, 2000). The following is some of the examples of rephrasing words or sentences given to the participants to further probe what they actually meant:

(To Teacher C): Just now you mentioned “task-based assessment”, can you talk a bit about your feelings? That is, when you are doing ‘task-based’, are your feelings mainly good? Why is it good? Or why is it no good, say when you encountered technical problem? Can you talk a bit more?

(To teacher B): Understood, er, that is what you were saying just now: there is formative assessment and summative assessment, right? So for these two types of assessment, how do you obtain the tools? Just now you have already shared that you have to specially design them, right?

(To Teacher H): So earlier you mentioned about ‘validity’, so do you think this assessment is valid?

(To Teacher I): But for your case, from what I gathered is feasible, because, for example you shared earlier about writing a first piece of composition, oh sorry, it should be an email, right? Email. So first piece I give you feedback and I give you feedback again on the second time. When it comes to the final submission, a mark will be given. So in other words, it has a process right?

Rephrasing the words or sentences by the participants also helped them to refocus on those important issues relevant to the research question. Although it was not easy, judgemental comments, both positive and negative were avoided by me during the interviews. The idea was to allow the participants to voice their experience of the phenomenon without many input and influences from the interviewer. However, maintaining a natural conversation during the interviews was also essential. If not, the participants might feel uneasy and even nervous as the researcher kept quiet all the time.

Last but not least, the Type 5 questions were used to conclude and end the interview. I would say “That is about all, do you have any other things to add or comment?” In this way, I provided the last opportunity for the participants to think and speak on any outstanding issues which had not been touched on earlier. This would complete the whole interview process.

3.5 Data Analysis

Data analysis can be considered as a systematic search for meaningful findings. It is a way of processing the qualitative data so that what has been done and learned can be communicated to others in an effective manner. A very crucial issue in phenomenography analysis is how the researcher analyses or accounts for the interview data. In this research, I followed a process (see Figure 3.5 below) which started from the interviews till the end when the “outcome space” is created. This step-by-step process of analysis outlined here is a method of demonstrating transparency of how I formulated the overarching themes from the interview data to the creation of the final “outcome space”. Once the interviews are conducted, transcriptions are carried out. After multiple reviews on the transcriptions, translations are done from Chinese to English, before coding (deductive and then inductive) could be carried out. Then I went into a vigorous process of reviewing and re-coding of data to arrive at the “categories of description” before finally creating the “outcome space”. Although presented as a linear, step-by-step process in the diagram, the data analysis was an iterative and reflexive process.

Interviewing and analysing participants’ experience of AT is the core of this phenomenographic research. The approach taken basically followed what Marton (1988, 1994) and Marton and Booth (1997) advocated earlier. The nine CL teachers’ interview data was audio-recorded and transcribed. During the initial stage of the data analyses, the interview contents were listened to and reviewed many times. A preliminary “categories of description” was conceptualised based on the analysed codes. See Appendix F to see how the conceptions are analysed and named. As I went on to review and re-analyse the data, many versions of the conceptions of AT were formed. See Appendix G.

From the interview data, the meanings were made from the teachers’ perspectives and not from the researcher’s point of view. As Svensson (1997) has explained, the fundamental assumptions about the objects of phenomenography have to do with the nature of conceptions. The ontological assumption of phenomenography is that the meaning of an experienced phenomenon is constructed by the “experiencer” and the phenomenon. As such, the conceptions signify the relationship between the CL teachers and the AT. In other words, the relational nature of phenomenography focuses on the relationships between the experienced phenomenon and the “experiencers”

(Dall’alba, 2000). Another ontological assumption of phenomenography is that the nature of conceptions is closely related to the nature of knowledge and thinking. Knowledge is assumed to be created through human thinking and human activity. Nevertheless, knowledge is also seen as to be closely related to the world or reality experienced by the participants. To this end, knowledge and conceptions are very closely related.

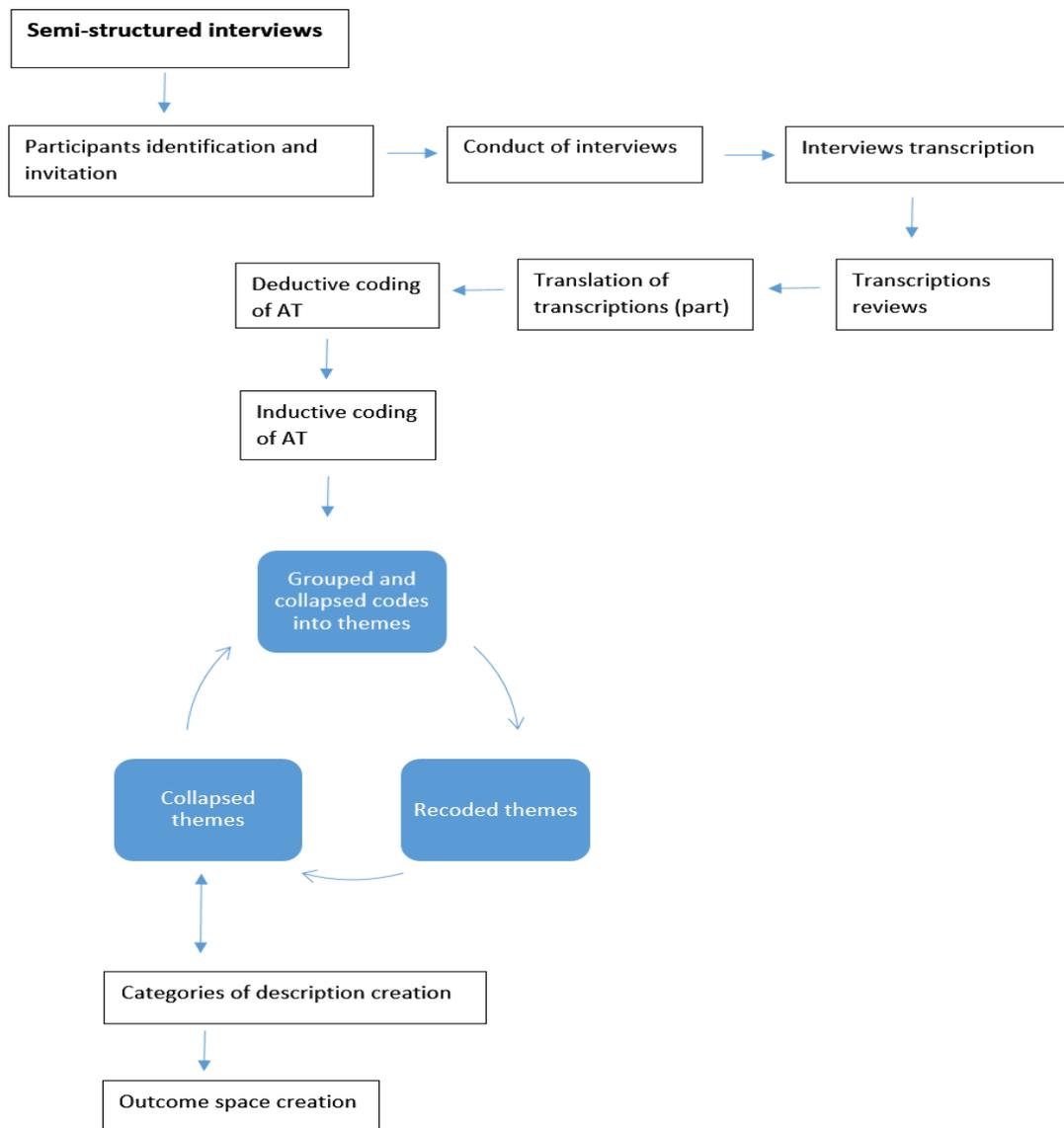


Figure 3.5: Process of phenomenographic data analysis

In this research, the lengths of the interviews ranged from 30 to 70 minutes. The average time of the interviews was 45 minutes. Transcriptions were verbatim and in

Chinese. I personally transcribed the first four participants' interviews to be familiar myself with the data but had chosen to leave the rest of the interviews to a transcriber due to time consideration. The transcriber had some experience in educational research and transcription. I also provided her with some basic training for how best to transcribe my interview data. All the transcriptions were reviewed and translated by me before they were advanced to the analysis stage.

At the beginning of the analysis stage, I first approached the transcripts using the deductive coding method because I have already got some ideas from personal experiences and past literature of the multiple considerations in teachers' assessment practices which they would experience and manage. The CL teachers' experience of AT obtained from the interviews were grouped since there were many ways of experiencing this phenomenon and each has not expressed the same experience in the same way. A preliminary idea of the possible ways in which the participants experienced AT was described based on the transcripts. Then I moved on to the inductive coding stage where key words, special terms, aspects or issues that the teachers frequently described, highlighted or paid attention to were carefully noted, grouped, collapsed and reviewed again.

On each of these review sessions, the reading of the transcripts was a new experience for me but I still read with the same focus as a whole. I initially used the "cut and paste" method which allowed me to extract particular key words, sentences or paragraphs and put them together in a separated summary section within the same transcript. However, I found that the coded categories did not come out strongly in meanings. As such, I realised that I needed to return to the beginning of the transcripts to start reading and coding again. By continuously going forward and backward, this analysis process allowed me to establish the preliminary "categories of description" more confidently. Despite this tedious and careful way of analysing the transcripts, I felt more confident each time a new version of the "categories" was produced. After a few weeks of reading and re-reading, grouping and re-grouping of the transcripts, I obtained four sets of AT with three distinct "categories" (see Appendix F). At this stage, these "categories" were not logically related by the dimension of variations, though it might seem to be somehow related at a first glance. They were indeed different from each other but I could not establish a relationship that is hierarchically related.

Then I began a major second round of reading each transcript's summarised segment, regardless of the initial the "categories". As I practiced this iterative and reflexive process of reading and reviewing of these summarised segments of the transcripts, I began to create new "categories" again. The main difference between this round of readings from the initial round was the focus. While the previous round was on summarising the themes of the interviews, the second round was on the relationships among the "categories". I focused on whether they could be understood in relation to each other. I would either change the "categories" or abandon the themes altogether.

It was only after I was satisfied with the relationship between different "categories" would I start to refocus on these "categories" to look for description and quotes. Progressively, I began to confirm a set of three "categories" that could be understood as a full progression (or nested hierarchy) in terms of the ways the nine CL teachers experience AT. The "categories" could be recognised by the critical aspects of the phenomenon (assessment purposes, reality and ideal gap, and demand from stakeholders) progressing from the first conception to the more advanced ones in a ranked fashion. When I could identify and explain the complexity of the conceptions from the first to the most advanced, I knew I had already established a structure of qualitatively different ways in which CL teachers experienced tension in assessment. These "categories" could exhibit the similarity in understanding among the transcripts assigned to each category and the differences between the "categories", a structure which Bowden (2000) had described.

This whole episode was important because phenomenography is essentially employed to describe the experienced phenomenon from the perspective of the participants and ultimately seeks to understand the different ways of experiencing AT and their interrelatedness. I aimed to analyse and understand at as deep a level as possible what has been said and what they actually meant. This analysis was to make sense of the particular expressions in terms of the individual, the group of participants, as well as their interrelated relationships.

3.6 Role of the researcher

The experience the researcher has as an educator would greatly influence how the research was conducted. I was a classroom teacher for 11 years and my assessment

practices were guided by how I conceptualised teaching and learning. How I conceptualised teaching and learning depended heavily on how I was being taught in school as a student and subsequently the experience I have as a teacher. However, teachers' behaviours are also often dictated by the teacher appraisal system (the Enhanced Performance Management System or EPMS) in Singapore. EPMS pits teacher against one another in the same substantive grades. They will compete to get a performance bonus or promotion based on a group of KPIs. Very often, what the ministry and school leaders will try to measure is how well the teachers provide the kind of training that will enable their students to do well in the national examinations. According to my experience, the EPMS has great implications and impact on teachers' conceptions of teaching and assessment as they attempt to "perform" based on the KPIs rather than their original conceptions of teaching and assessment.

Being a former CL classroom teacher teaching Secondary 1 to 5 and a Head-of-Department (HOD) of MTL in a Singapore school, the issues and challenges CL teachers encountered are not strange to me. Together with my current attachment with the MTL Branch, Curriculum Planning and Development Division (CPDD) of the MOE in Singapore to develop, plan, revise and review the curricular materials for the 2011 newly implemented CL syllabus, I have the opportunity to observe the new CL curriculum development and its implementation. Working closely with the colleagues from the Singapore Examinations and Assessment Board (SEAB) on the revised assessment modes and specifications of the 2011 newly implemented CL syllabus, I also had the opportunity to see how the SA is being conceptualised, designed, field-tested and finally implemented. In addition, being one of the main members from CPDD to organise the annual CL school's zonal dialogue sessions with classroom teachers and HODs, I could also anticipate and feel the frustrations, tensions and challenges they might face as a result of the "newness" and controversies within the 2011 CL curriculum and assessment, especially the content and the paper formats of the CL GCE 'O' Level examination. All these experiences have allowed me to see from the teacher participants' perspectives in this research. As a result, my "insider" knowledge will also be shaped by them.

3.7 Validity, reliability and “bracketing”

In research, validity refers broadly to the “goodness” or “soundness” of a study. Depending on the research methodologies and perspectives guiding the research, there are multiple ways to conceptualize validity. As such, the validity of qualitative research is often judged by an “individualized contextual manner” rather than by a “broadly applicable standards and criteria” (Miller, 2008, p.909). Indeed, in the case of phenomenographic research, according to Dahlin (1999):

the validity is based on three factors. The first is the logic of the system of categories emerging from the analysis. The categories must be logically separate and exclusive. The second factor is the correspondence between the results and what is known from previous studies in the field. Finally, the plausibility of the categories may be considered, i.e the extent they are recognisable as representing actual or possible human experiences (p.195).

In this research, the “categories” were separated logically in the sense that there was a systematic way to transcribe and sort out the participants’ conceptions or experiences of assessment into various “categories”. As explained earlier, the interview transcripts were read with openness but subsequently focused on a particular aspect or criteria where references were taken from past studies on teachers’ experience of AT after multiple readings. During this process, there were constant sorting and re-sorting of transcript data, as well as ongoing comparisons between the “categories”, paying special attention to the qualitative similarities and differences between the “categories”. There were many re-readings, revisions and thus multiple versions of the categories of description. This was to ensure that this study explored all the possible perspectives from the interview data. Together with the data obtained from the other sources, the experience of the participants was described more accurately.

The final confirmed conceptions were recognisable because the critical aspects of each different category of description were carefully identified and presented in a concrete and hierarchical-related form. These qualitatively different ways of experiencing AT were the most important findings in this research. Once the concrete

and hierarchical-related conceptions or experiences of assessment of the participants were identified, a more detailed description of their typical personality or characteristic of each conception is presented and discussed.

Many are aware that qualitative researchers do not view reliability the same way as quantitative researchers. According to Bogdan and Biklen (2007), “qualitative researchers tend to view reliability as a fit between what they record as data and what actually occurs in the setting under study, rather than the literal consistency across different observations” (p. 40). This implies that two different researchers studying a same phenomenon may produce different data and findings, yet both studies could be considered reliable. Marton (1986) highlights two issues of reliability (or replicability) in phenomenographic research. The first issue is whether other researchers could reach the same “categories” as the original study. The second issue is whether the other researchers can identify the conceptions found by the original study through the “categories”, something which is termed “interjudge reliability” (Marton, 1986; Sandbergh; 1997; Saljo, 1988). However, Sandbergh (1997) argues that interjudge reliability is problematic because it creates a theoretical and methodological inconsistency within phenomenographic study and therefore should be abandoned. He then goes on to suggest five steps or procedures for phenomenographic researchers to enter “phenomenological reduction” so that “reliability as interpretative awareness” can be maintained to overcome the problem of establishing reliability of phenomenographic research (p. 211). In this research, one of the many efforts to address the reliability issue was to conduct pilot interviews. Researchers can enhance the reliability of their interviews through piloting their interviews (Silverman, 1993). In these pilot interviews, the attention was given to the clarity and structuring of the questions, as well as the time needed for the actual interviews.

When analysing phenomenographic data, the intersubjective agreement forms the basis for the reliability of the categories of description found. There are two types of “intersubjective agreement” with phenomenographic interview data. The “dialogic reliability check” could be conducted where “agreement between researchers is reached through discussion and mutual critique of the data and of each researcher’s interpretive hypothesis”. The second type is the “coder reliability check”, where two or more researchers ‘independently code all or a sample of interview transcripts and

compare categorisations' (Åkerlind, 2005, p.331). Similarly, Bowden (2000) suggested all interview transcripts to go through "rigorous phenomenographic analysis". This means one of the research team takes up the initial task of reading all the transcripts and devises "a draft set of categories description" drawn from them. This team member will review these transcripts one more time and then make allocation of transcripts to categories. On the other hand, another team member will also carry out the same transcript reviews and make allocation of transcripts to categories independently. Finally, these two members will then compare their two allocations for similarities and differences. When there are disagreements between the two members, reference will be made to the transcripts as the only evidence for further discussions and allocations. While this research was conducted independently without other team members, my supervisor, experts and colleagues in the field of phenomenography were invited to "cross-check" my analysis and categorisations. Earlier, Bowden (2000), Prosser (2000) and Trigwell (2000) has advocated a "dialogic dependability check" in their studies to ensure trustworthiness. This is also similar to the idea of "coder dependability check" during the development of the categories of description by Collier-Reed, Ingerman and Berglund (2009). Allowing different members or experts to independently to check on the coded interview transcripts and compare their versions of categorisations will definitely enhance the validity, reliability and thus the trustworthiness of this phenomenographic research on AT.

Taking into considerations of the scale and resources of this research, dialogic reliability check was adopted instead of the coder reliability check. I first analysed and then presented the categories of description to my supervisor. Both parties usually started off with a general discussion on the categories of description. When there was a need, I would argue and defend the findings against the questions and challenges posted by the supervisor. This process was carried out several times until both parties were satisfied with the decided categories of description. During the process, I also wished to know if others could understand the "categories of description" based on the selected transcripts. I took opportunities at various platforms to interact with scholars and researchers so that I could obtain feedback on my findings. Over the past few years, I presented my findings at Graduate Student Day in both 2015 and 2016, Assessment Graduate Student Seminar 2016, Curriculum, Teaching and Learning (CTL) Academic

Group Doctor in Education (EdD) seminar 2016. These events were organised by the National Institute of Education (NIE). My most-recent presentation of this research was at the 2018 conference of the European Association for Research on Learning and Instruction (EARLI) Special Interest Group (SIG 9) on Phenomenography and Variation Theory at the University of Birmingham, United Kingdom. The SIG 9 participants commented on my outcome space and made minor suggestions on the adjective I used to describe Conception C. In summary, at these platforms, I consulted many individuals on my findings, particularly on the categories of description and outcome space. The main audiences were academic researchers and scholars, Master and Doctorate students, and a few phenomenographic researchers. These interactions with experts who are familiar with phenomenography and the feedback provided by them has proven to be useful as they kept me thinking and revising the categories of description.

Another very crucial issue in phenomenographic research analysis is how the researcher accounts for the interview data. To a large extent, the accounting of the interview data concerns the validity and reliability of the research and “bracketing” was a critical aspect for addressing these two issues. Marton (1994) and Ashworth (1999) have discussed about the importance of “bracketing and posited that there is a need for researcher to carry out “the suspension of presuppositions” (the “epoche or bracketing”) when analyzing data. The task is actually “to set aside theories, research presuppositions, ready-made interpretations etc., in order to reveal engaged, lived experience” (p. 707).

On the same note, Tan (2010) pointed out that some of the more prominent presuppositions the phenomenographic researchers should “bracket” includes (a) personal knowledge and beliefs; (b) assumptions of the nature of the subjects’ experience; and (c) excessive concern over why certain experiences had occurred (p. 44). Indeed, my personal knowledge and belief about CL teaching, learning and assessment in Singapore, as well as my contact and understanding about CL teachers would affect how I frame the interview questions and do the interpretations. In this research, “bracketing” was practised throughout the data collection phase to the report writing phase. The conception of the researchers about a given phenomenon is not the focus of the phenomenographic research, instead the focus should be about the conceptions of the participants have on AT. As such, keeping an open mind at all stages of the study and not taking the researcher’s perspective will help in the “bracketing” process. For this

reason, the interpretive procedures I had followed in this research became paramount importance. I needed to account and communicate to the readers and experts of phenomenography of the steps and efforts taken to analyse the transcripts, and how I arrived at the categories of description, as well as the final outcome space.

3.8 Summary

In this chapter, the methodologies employed by past studies on AT were reviewed, and it was concluded that they have not adequately captured the qualitatively different ways in which teachers experience AT. In other words, they have not sufficiently investigated the collective difference or variation in which teachers describe their real encounters with AT. Subsequently, this chapter went on to present the unique contributions of phenomenography to AT research and how the gap left by past studies may be closed by this approach. Particularly, this chapter has argued that phenomenography is a suitable methodology to investigate teachers' experience of AT, given its complexity. Most importantly, the theoretical, analytical and pedagogical benefits of knowing "teachers' ways of experiencing AT" for this research was also highlighted. This is in line with Marton and Booth (1997) and Tan's (2010, 2013) arguments on why phenomenography is an appropriate research approach for exploring teachers' qualitative different ways of experiencing a phenomenon. Phenomenography's essence is to explore the "variation in ways of experiencing things" (Marton & Booth, 1997, p.110). It aims to describe something that is experienced by an individual, from the perspective of the individual.

For the rest of this chapter, I have described how the data sources were collected and analysed, what the role of the researcher was and how the validity, reliability and "bracketing" issues could be taken care of. In the next chapter, I described in details the three identified conceptions of AT that formed the outcome space of this research.

CHAPTER 4

FINDINGS

4.1 Introduction

The purpose of this research is to describe the qualitatively different ways in which CL teachers experience AT. This chapter describes the variation in the conceptualisations of AT encountered by CL teachers in Singapore and the various ways in which they attempt to address, manage or deal with these tensions. To be more specific, this research addresses the main question: **What are teachers' conceptions of AT?** I had defined AT as "teachers' feelings of struggle with competing considerations in their assessment practices". This definition of the phenomenon of teachers' experience of AT could be logically reduced into the 'What' aspect and the 'How' aspect of the phenomenon:

- (1) What assessment considerations do teachers focus on?
- (2) How do teachers deal with these competing considerations?

Introduced in Chapter 1, tension as a concept is more commonly discussed in physics and it refers to "the act of stretching or straining" or "the state of being stretched or strained". In these cases, stretch and strain have very similar meaning of a force. When "tension force" is being discussed, it will invariably involve the concept of "Newton's Laws of Motion" for a better understanding of the underlying fundamentals.

Newton's first law states that an object that is at rest, or not moving, will stay at rest unless a force acts on it. A moving object will continue moving at the same speed and *in the same direction unless a force* acts on it. This means an object fights any change in its motion. This fighting is called inertia (Chiang, 2011). This law is also known as the law of inertia. This is a force that resists a change in an object's motion. In this research, I found that the teachers' assessment experiences, too, are subjected to these natural laws that closely resemble the laws of motion. One possibility is that the teachers may turn out to be very passive in their assessment practice and this may be understood as the law of the teachers' inertia towards tension in assessment. That is the tendency of the teachers, having once established a comfortable assessment practice, to continue to stay in that "comfortable zone" unless acted on by a greater force. When the teachers'

focus is on fighting against any change in the direction of the motion, it will always stay passive and inactive.

Newton's second law of motion says that acceleration is dependent on the mass of an object. The net force is the sum of all the forces acting on the mass (Zimba, 2009). In general, the greater the mass of the object, the greater the amount of force needed to mobilise the object. As this force is stored within the object, it is also said to have potential energy, but once a force is exerted or responded on the object, the energy is said to have been changed to kinetic energy. Drawing on Newton's second law, I observed another group of teachers whose experience of AT is more responsive. It can be understood as another law of human nature towards their experience in AT. There seems to be a tendency of the teachers, having the capacity to face challenges and changes, to wait for the opportunity to show response and exert some force to adjust the course of assessment practice and move to a different direction. When the teachers' focus is on what they can possibly do, the energy or force behind them will be very responsive. This is like the water behind a dam that stores great potential energy. When the water is released from the dam, the potential energy is changed into kinetic energy.

"Kinetic" comes from the Greek word for motion (kinesis). Kinetic energy occurs from motion and the ability of the object in motion to do work. One of the three forms of kinetic energy is called "translational" and it is generated when an object moves from one position to another. In other words, kinetic energy can be transferred to other object(s) or forms (Lawrence, 2016). Taking on the concept of kinetic energy, I observed the most sophisticated way of experiencing AT. It can be understood as another law of teachers' forward-looking and forward-going attitude when they experienced AT. The tendency of teachers, having the foresight to keep improving their assessment practice; to constantly reflect on their current assessment practices and re-conceptualise assessment issues so that they can keep moving forward with a moral purpose.

In brief, Newton's law of motion allows readers to understand that motion is caused by forces. It can get things moving and change the current state of things. However, we also learn that forces do not always make things move. For example, a huge building may have massive forces acting on it, but it does not go anywhere. In addition, the forces produced could possibly shift the direction in which something is moving or even change in its form. Understanding teachers' experience of AT in terms

of physics provides readers with a good starting point to learn about the various ways in which teachers may experience AT in practice and the potential responses they may have toward the same tension in assessment.

In this research, the interviews were conducted and transcribed in Mandarin Chinese so that participants' original meanings and their natural professional lexicon can be best preserved. The data analysis was done with a mix of Chinese and English, with the final results reported in English (including the translated transcripts). This is one of the features of this research, which seemingly to have compromised the quality of the findings because of the gap created by this change in language. While the change between Chinese and English may seem to have, to certain extent, undermined participants' immediate and rich account of their assessment experience, it is a naturally inherent tension that is often experienced by the subject of this research in the Singapore context.

I have attempted to use the physics concept in the following section to specially explain CL teachers' experience of AT. The main aim is to inform readers about the nature of "tension". In particular, how the teachers' experience and awareness of AT advanced from the "basic" to the more "sophisticated" understanding is presented in a more accessible manner.

4.2 The outcome space

The outcomes of this research on teachers' experience of AT may be understood in three related and distinct ways. The outcome space is made up by the three conceptions of AT. Each of these three conceptions are represented with an adjective to describe teachers' qualitatively different ways of experiencing AT- passive, responsive and propulsive. Each conception is derived from the phenomenographic analysis and translated from Chinese to English. Passive was translated from the Chinese word "消极", which literally means "negative" or "passive". In this conception, AT is passively contained by current limitation. Responsive was translated from the Chinese word "积极", which literally means "positive" or "active". In this conception, AT is responsive to imminent challenges or opportunities. Propulsive was translated from the Chinese word "推进", which literally means "driving" or "propelling". In this conception, AT is

propulsive to energise teachers to go beyond current assessment affairs. The attributes of each conception are offered through how the teachers experienced and managed the experienced AT. Thus, the passive conception would describe what these groups of teachers understands and does with tension in assessment. Likewise, the responsive and the propulsive conception depict the other variation the teachers experience AT in their distinct ways.

In phenomenography, the qualitatively different ways in which teachers experience AT are denoted as “conceptions” and is represented by “categories of description”. The three conceptions of experiencing AT in this research are found to be the passive, the responsive and the propulsive. Correspondingly, they assume the role of compliant executor, rational practitioner and advanced re-conceptualiser in their assessment practices. The entire complex of these conceptions encompassing all the different ways of experiencing this phenomenon is the outcome space (Marton & Booth, 1997). While many tensions were identified at the preliminary analysis from the interview data, the AT discussed in the following section is determined by the extent and depth of discussion during the interviews. To be more precise, the following conceptions of AT elicited the largest amount of conversations during the interviews. These are the qualitatively different ways of CL teachers’ experience of AT. Typically, a way of experiencing AT consists of CL teachers’ awareness of the critical aspects of the phenomenon and the structure of these aspects in relation to each other. Their different ways of experiencing AT varied because of the number and the structure of the critical aspects that are experienced by them. The table below summarized the three different ways of experiencing AT, at a more general level.

	Three different ways of experiencing AT		
	Conception A: Passive (消极的)	Conception B: Responsive (积极的)	Conception C: Propulsive (推进的)
The ‘What’ aspect of the phenomenon	*Limiting teachers’ assessment practices to stay status quo	*Prompting teachers to leverage on the imminent opportunities	*Propelling teachers to go beyond current limitation and imminent

		to change their assessment practices	opportunities in assessment practice and to use assessment for greater purposes of education
The 'How' aspect of the phenomenon	Compliant executor (服从的执行者) *Confined teachers to stay with status quo and are not willing to change any assessment practice *AT is a restrictive force	Rational practitioner (理性的实践者) *Made teachers respond to the challenges by making some changes in assessment practice *AT is a useful force	Advanced Re-conceptualiser (前卫的概念重建者) *Enabled teachers to think deeper into assessment in the long run *AT is a constructive and useful force

Table 4.2 Teachers' qualitatively different ways of experiencing AT

These conceptions of AT are characterized by variation along some key dimensions (or critical components) in their categories of description, which serve both to link and separate CL teachers' different ways of experiencing AT. These dimensions not only define the structure of the outcome space but also represent the variation in conceptualizing AT. At the general level, the variation is defined by whether the focal awareness of the teachers' experience is on current limitations, imminent current opportunities (students' short-term learning) or beyond current limitations and opportunities (students' long-term learning). These are the 'What' aspect of the phenomenon. Each of these distinctive experiences is denoted by the conception and is accompanied by an "adjective" to describe and summarise the nature of the tension experienced. In fact, I have used these "adjective" as an analysis to capture more of this individual richness. This 'What' aspect of teachers' experience of the phenomenon is rationally related to the 'How' aspect of it in that what teachers conceptualise AT to be

will affect 'how' they react or deal with this tension. For example, if a teacher experienced AT to be something negative, her or his response to this would hardly be very positive as well. My above observations turned out to be very intriguing and complex. The actual scenarios, both in the context of physics and teachers' experiences of AT, were highly complicated.

Conception A (Passive)

It has never been easy to change teachers' assessment practice. Many would think of the law of inertia as any object that is at rest and not moving at all. I find that the reality of inertia in the teachers' experience of AT in this conception to be rather different. Newton's first law of motion describes not only the energy acting on a static object. Instead, the teachers are in fact moving quickly and inescapably along a teaching track that is driven by powerful education system and its stakeholders. Analysing the teachers' experience of AT in this dynamic perspective changes our understanding of what it takes to bring change to teachers' assessment practice. I observed that the teachers may not be stuck, as so many referred to themselves when they experience tension in assessment. Instead, Conception A depicts the conceptualisation of tension in assessment as current limitations, which naturally chained the teachers to stay passive and do only what is within their immediate control in assessment practice. This tension confines teachers to stay with the status quo and makes them unwilling to change any assessment practice. In this conception, AT is experienced as a restrictive force that prevents CL teachers from making assessment adjustment.

The key attributes of this conception include a focus on maintaining the status quo and pressure to conform to the current limitation in their assessment practice. Teachers' assessment focus is on students' short-term learning. Logically, one of the attributes of the teachers who conceptualize a passive conception of AT is that of a compliant executor. This is the 'How' aspect of the phenomenon. Typically, such teachers always lament that their assessment actions are being dictated by other external forces and they have no choice but to follow the crowd. There seems to be a lack of ownership in assessment but it could also be that they are "protected" by school leaders to maintain such passive practice. In particular, these teachers deal with AT by trying to preserve the present state of affairs and not making any changes. Teachers

with this conception will naturally respond to this phenomenon by not trying new ideas in assessment.

The passive conception of AT may be understood as a continuous focus on the current limitations in the teachers' assessment practice. Teachers in this conception usually feel helpless to change the course of their assessment practice. To them, as much as they want to or have tried to, they just cannot seem to alter the reality in assessment. Newton's first law of inertia is so apt to explain this phenomenon. If teachers are going to change anything in their assessment practice, they need to apply forces that are greater than the forces currently controlling the assessment situations. These teachers typically respond or deal with the phenomenon by not making any changes and stay with the status quo in their assessment practices. Indeed, compliant has the meaning of "lacking in energy or will" and an executor is an enforcer who merely carries out orders without much thought. Put together, compliant executors are passive and do not perform any additional duty outside their designated scope of teaching and assessment tasks. They tend to follow strictly the basic assessment job scope without thinking about other possibilities. Their assessment focus is on students' short-term learning and achievement. Such teachers also tend to stick to teaching and examination requirement which are mostly standardised. When they experience AT, they would deal with it by not responding.

Conception B (Responsive)

When there are multi-directional "tug-of-war" battles from various forces, the resultant force depends on who finally wins the "tug". So in reality, AT is moving at different speeds (and thus powers) propelled by multiple forces in teaching. Depending on the amount of forces from the teachers, the resultant force can have the great potential to counteract those forces which are currently making them struggle. Very often, because of the great forces that are already challenging the teachers' assessment practice, even greater forces need to be exerted if there are going to be significant changes. While Conception A depicts CL teachers' focal awareness on current limitations when they experienced the phenomenon, Conception B has a broader focus as it shifts to experiencing AT as potential and imminent opportunities to improve a current reality in the teachers' assessment practice. This conception is more sophisticated than the

previous one as it surpasses the focus of current limitations and conceptualises AT as opportunities for teachers to counteract and make changes in their assessment practice. However, teachers' overall assessment focus is still on students' short-term learning. The key attributes of this conception include the perception of AT as a force for them to take ownership in assessment practice. It enables teachers to face the challenges in assessment by being responsive and making some changes in their practices. In this conception, AT is experienced as a responsive force that invites teachers to take ownership of assessment and make necessary changes to improve their practice. The intention for such conceptualisation and subsequently their response to the phenomenon is very clear. It is to improve students' learning so that they can do well in the national examination.

Unlike the passive conception, a responsive conception of AT differs in that the teachers' focus is on imminent opportunities, rather than on current limitations, though both are considered with respect to students' short-term learning. When faced with an assessment challenge, the teachers in this conception are able to take initiative to make changes in their assessment practice. As such, exploring other possibilities and taking ownership in assessment seems to be the attribute of this conception. This responsive conception posits that teachers need to actively respond and to adjust their assessment practice when they are being challenged. For the 'How' aspect of the phenomenon, one of the most prominent attributes of teachers with a responsive conception of AT is that of a rational practitioner. While being rational means a person's action is guided by some sound reasons, a practitioner portrays a sense of professionalism. Such teachers are willing to take up ownership and are empowered to take actions when there are assessment issues. They deal with AT by adjusting and changing their current assessment practices. Unlike the compliant executer in the previous conception, these rational practitioners are more proactive and professional to ensure that they manage their assessment problems to good effect within their area of control for an intended purpose.

Conception C (Propulsive)

Conception C has the most sophisticated focal awareness in the outcome space and is differentiated from the previous two conceptions in that the teachers' focus is

now beyond the current limitations and imminent opportunities. Their assessment focus is on students' long-term learning that is often beyond schools. In particular, they do not focus on students' short-term needs in assessment. Teachers with this conception of AT view the phenomenon as some forces which would propel them to look forward, reflect and even re-conceptualise assessment issues. The tension they experienced also enables them to think deeper about their teaching and assessment practices in the long run. Teachers with a propulsive conception typically are not afraid of the multi-directional "tug-of-war" battle from various forces. They refuse to be dictated to by these forces and seek to overcome existing difficulties in assessment practice by looking at the bigger picture of the MTL education in Singapore. For example, in Singapore, CL as a subject in secondary education is not a compulsory subject for admission into higher education, though there is a minimum threshold that every student must meet. That propels some teachers to think more deeply about the objectives of CL teaching, learning, as well as assessment in the longer run. If the CL national examination result is not mandatory for placement purpose of higher education (junior colleges and polytechnics), some teachers may believe that they have to reconsider the functions and impacts of their classroom assessment, especially in the long run.

Unlike Conception A and B, which tend to focus on current or imminent happenings, Conception C tends to project the teachers' focus on the impact of assessment in the longer term or students' future learning. For such teachers, AT is understood as an important reminder to reflect upon their teaching and assessment. Are they preparing CL students just for national examinations, or are they preparing them for lifelong learning? What is the real meaning of CL education? These are some of the typical questions that teachers in the propulsive conception of AT often ask themselves. As for the 'How' aspect of the phenomenon, one of the distinctive characteristics of teachers with a propulsive conception of AT is that they will initiate changes, both at the philosophical and practical domains of their assessment undertaking. They are highly confident that their way of thinking can guide students into better learning experience and ultimately benefit from their language teaching and assessment.

This AT is constructive and useful because it allows teachers to reflect on the functions and impacts of their teaching and assessment practices to their students in

the long run, often beyond schooling. Teachers with a propulsive conception of AT actively consider revamping current assessment practices and constantly ask critical questions of the beliefs and fundamentals of these practices. They are no longer contented with what is covered in the national assessment syllabus, as they are more interested to use assessment to improve their students' results and beyond. They do not privilege SA (such as the pervasive use of SA in ranking and placement) like most of the teachers. Instead, they emphasise students' lifelong or future learning. Teachers in this propulsive conception usually deal with AT with much confidence. The quality of such a "re-conceptualiser" is very sophisticated because it takes a lot of more awareness for them to pursue what they feel to be the best for their students' CL learning and assessment. They are grounded in the belief of the betterment in the teaching and assessment of the CL. This quality is similar to what Michael Fullan (2001, p.13-14) describes as a person who is filled with "moral purpose". They have a bigger vision than others and are deeply passionate about improving the current state of CL teaching and assessment.

From the above description and discussion, I argue that this hierarchical feature of the analysis served phenomenographic research very well as the qualitatively different ways of experiencing a particular phenomenon are usually presented in a nested hierarchy. Essentially, a more sophisticated conception should encompass a greater number of themes experienced in AT. While the above summary showed how CL teachers experienced AT at the collective level via the three conceptions, the following sections describe and analyse the exact AT they encountered in their teaching in greater depth. To summarize, the AT experienced by the CL teachers can be conceptualized and understood as assessment: (1) purposes; (2) reality and ideal gap and (3) demand from stakeholders. These findings provide a structured mapping of the various ways in which tension in assessment is experienced by CL teachers and would give further authentic evidence of the current challenges and opportunities in their assessment practice. They would also help teachers discover the types of AT, struggle and dilemma they currently encounter in practice, which they might not have been aware of previously.

4.3 Qualitative difference in teachers' experience of AT

This section focuses on more detailed descriptions on the 'What' and 'How' aspect of teachers' experience of AT. It presents the variation in which CL teachers experience and conceptualise tension in assessment. This variation can be understood in terms of a few dimensions. These dimensions of variation, as well as the attributes of the three conceptions of AT are summarised in Table 4.3 below.

A T	Dimensions of variation		Three different ways of experiencing AT		
	Understood AT in terms of	Distinctive AT experienced	Conception A: Passive (消极的)	Conception B: Responsive (积极的)	Conception C: Propulsive (推进的)
1	Purposes	Assessment for learning vs national examination (formative assessment vs summative assessment)	*Do not see the purpose of assessment in teaching *Focused on accountability in assessment *Summative assessment-oriented (unbalance) *AfL has no place in preparation for national exam	*See the formative and summative purpose of assessment as complementary *Focused on the quality instead of the balance in assessment *AfL has a place for both national exam preparation and improve learning	*See assessment purpose beyond just FA or SA *Focused on student-centred assessment *Assessment is not only to serve national exam but has its own place and purposes

2	Reality and ideal gap	Current assessment practice vs teachers' ideal assessment	<p>*Focused on national exam only</p> <p>*Focused on only forms of language assessment: the 'individual language skills' (listening, speaking, reading and writing)</p> <p>*See quality of assessment as: validity only</p>	<p>*Focused on both national exam and classroom assessment</p> <p>*Focused on forms and functionality of language assessment</p> <p>*See quality of assessment as: validity and authenticity</p>	<p>*Focused on all types of assessment</p> <p>*Focused beyond forms and functionality of language assessment</p> <p>*See quality of assessment as: beyond validity and authenticity</p> <p>*Re-conceptualise traditional language assessment and think beyond students' current CL learning experience</p>
3	Demands of stakeholders	Stakeholders' demand vs teachers' decision making in assessment	*Teachers' professional decision making fully compromised	*Teachers' professional decision making partially compromised	*Beyond answering to the assessment demand of stakeholders

			*Not much teacher empowerment	*Some teacher empowerment	*Teacher self-empowerment
--	--	--	-------------------------------	---------------------------	---------------------------

Table 4.3: The three conceptions of AT

4.3.1 Conceptions of AT in terms of purposes

In my practice context, CL teachers experienced the tension in carrying out AfL while preparing students for national high-stake examination. This conception of AT can be understood in terms of assessment purposes. Their experience of AT generated the most conversation from the participants who, in their years of teaching, have to address the AfL issues that included paradigm shift, time, amount, resources, materials, contents, syllabus coverage and at the same time prepare their students well for the GCE O-level examination. This is a national high-stakes examination because it is taken by students at the end of their fourth or fifth year in secondary school and their results will be used for admission to higher education (junior colleges and polytechnics). The variation in their experience of AT generated three categories of description that represent the increasingly complex ways of understanding the struggle teachers went through with AfL in terms of its purposes, against the backdrop of preparing their students well for the national examination because of accountability. Such teachers recognise the multiple purposes an assessment has to serve and most teachers find that these purposes may not always be compatible with one another, especially when they are pressured to produce good results in the national examination.

Category 1 (Passive): Teachers are limited to practicing only summative assessment because of accountability

Category 1 depicts the teachers' experience of AT as one that limits them to practicing only SA due to teaching accountability. Their experience in this category is very narrow because they basically do not consider the other purposes of assessment in their teaching. Teachers attribute their experience to the current limitation of not being able to practice AfL in the classroom. Their assessment practice is not within their sphere of control and thus is being conceptualised as "passive"- their action is not going to get

them anywhere. Teachers assume that there is only one purpose in assessment and that is to prepare their students for the national high-stake examination, which is very short-term in focus. Indeed, one of the attributes of teachers who possess a passive conception of AT is that of a compliant executor. Such teachers respond to AT by staying with the status quo and do what they are required or asked to do in teaching and assessment. They are passive and thus usually wait for others to take the first move and hardly make changes to their teaching and assessment practices, as exemplified by the teacher who gave the response below. It is to note that each extract is translated from the Mandarin Chinese into English. All the original extracts are found in Appendix H.

Extract 1:

*Teacher C: I never thought of formative or summative assessment. I don't really understand these terms. **Actually I think my practice is more 'exam-oriented'. Because I need to be accountable for students' parents.** That is when it comes to exams, school, students will need to know these exam scope and requirement and so we need to teach them beforehand, that is to prepare them well for exams. Therefore, I give them these (assignment). Yes, yes that's why I every time, that is these assessments, generally are designed according to the exam, directly related. Yes, I will not, I I seldom do anything beyond exams, those task-based ones.*

The teacher is candid in sharing that she does not know what FA and SA are. She just “executes” the assessment according to the national examination specifications and requirement because she is accountable for the parents. This particular AT had limited her from practising other types of assessment in her teaching. Teachers in this conception deal with the phenomenon by not doing anything that is beyond the examination requirement.

In this case, the teacher did not think about the various purposes of assessment in her classroom. The AT she experienced from assessment accountability in her teaching had greatly limited them to stay with the status quo and followed what is mandatory in assessment. This is very typical of a Singapore secondary school where

students' assessment results are recorded into CA and Semestral Assessment. CA mainly include class tests, journal entries, quizzes and other exercises formally given by teachers in class. At the same time, Semestral Assessment consists of common tests and all school examinations. Both CA and Semestral Assessment scores would be recorded into the final computation of students' academic performance. Indeed, all these assessments added up to a large amount of work and would give pressure to teachers to forgo AfL tasks that needed much time and effort to prepare and implement.

In brief, this category denotes that the teachers perceive that they do not have much control over the current limitations on assessment and their main focus is on how to prepare their students well for the national high-stake examination in due course. They do not seem to have many choices in their assessment practices and consequently, AfL has no place in their teaching. This passive conception posits that all assessment practices should be used to serve the national examinations. To this end, their assessment focus is on students' short-term learning. Rationally, one of the common traits of teachers who conceptualize a passive conception of AT is that of a compliant executor. They typically deal with the phenomenon by not making any changes and stay with the status quo in their teaching and assessment practices. Such teachers tend to stick to their old practices in assessment.

Category 2 (Responsive): Teachers consider the different roles of assessment

This category increases in complexity from the previous one as the teachers' experience of AT is now shifted away from the current limitations to imminent opportunities. These include the call to use AfL and its related components, materials or resources for teachers to prepare their students for national examinations. The main attributes in this category are the teachers' considerations on the different roles and purposes of assessment, against the backdrop of preparing their students well for national examination. They see the formative and summative purpose of assessment as complementary instead of an exclusive binary. While the force behind the previous category was passive, Category 2 has advanced to responsive because the teachers in this category believe that both FA and SA are here to serve students' language learning. Although both have different purposes, the ultimate intended goal is to obtain good

results in the national examination. They actively embrace the use of FA (or AfL). Below is a good example:

Extract 2:

*Teacher 1: Er, assessment for learning, I think teachers should accept it with a positive thinking. In other words, we should, firstly, not waste time and secondly, we should not be asked to see the effect immediately, I want to see the teaching effect. How do we not face with the stress of exams? We must overcome this challenge, **use this assessment for learning to let your teaching to be more fantastic, to be more effective and at the end their national exams will not be poor.** This is my conclusion; they will not score poorly. It allows me to prepare my students for exams more confidently. I think that's really a way to helping me, my students to do well in the exams.*

Apparently, the teacher saw the purpose of FA in her language teaching as an imminent opportunity to get her students ready for the national examination. She believes that both FA and SA play a complementary role in her students' learning. Instead of just focusing on SA as in the previous category, the teacher who understand AT as responsive assume that AfL has a place for both national examination preparations, as well as to improve students' language learning. In addition, the teacher in this category focus more on the quality of assessment, instead of the limited functions of it.

To sum up, this category is more sophisticated than Category 1 in that the teachers are more interested and responsive to the tension bought about by national examination. They would capitalise on these "imminent opportunities" to meet their teaching and assessment objectives. In contrast to the previous category, the teachers with a responsive conception of AT will go beyond the status quo when they encounter challenges and tensions in assessment. Such teachers would endeavour to find ways or make adjustments to their current teaching and assessment. Unlike the compliant executor, the rational practitioner of the responsive conception would attempt to solve their assessment problems within their scope of professional knowledge and control.

Ultimately, their intended aim is to also prepare their students well for the national examination, which is still very short-term in focus.

Category 3 (Propulsive): Teachers longing to impact students' future learning while preparing students for summative assessment

This category is distinct from the previous ones in that the AT teachers experienced is more complicated as they now shifted away the focus beyond current limitations and imminent opportunities. In fact, this category is the highest level in this AT for teachers exhibit a paradigm shift in what they traditionally believe classroom teaching and assessment should entail. To this end, this conception is considered as “propulsive” because the teachers want their current teaching and assessment to have an impact on their students' future learning while preparing them for SA. In short, this conception requires a more fundamental change in the way teachers think of the direct impact their assessment can bring to their students. Comparatively, these teachers are more passionate about improving the teaching and assessment of the CL. However, their vision is not only on the national examination because they think it is too narrow a focus. Their assessment focus is on students' long-term learning that is beyond schools. To this end, the assessment in their classrooms has its own place and purposes. This vision is much wider and further as compared with the previous categories. In fact, the teachers in this category are able to decide the worth and relevance of their assessment practice and experience. They also often re-examine the meanings behind these values and are willing to re-conceptualize their belief and practice for the purposes of assessment in the long run. The extract below shown an example of such a re-conceptualizer:

Extract 3:

*Teacher 1: My challenge in assessment is how to get at least 50% distinction in exams. Yes, to do well in the exams. **At the end of the day, the student can use Chinese as a career or profession, or related to language, you know or be a DJ in a radio station.** That is big, that is my biggest challenge. I have not achieved that yet. For example, Desmond Koh. For example, Huang Shuangxi and so forth. The student can use Chinese as a tool for their work in future. They*

have to be among the best in Chinese. Then I think it will be a big success in using assessment.

The AT of working towards good results in the national examination has transformed into a propulsive force to allow the teacher to project her vision of using assessment to motivate her students to maintain a high standard of CL proficiency. This proficiency will eventually help the students to apply the learned language in their future professions. This category is distinct from the previous one in that the teachers' focus is on students' future use of the learned language, on top of doing well in the national examination. Accordingly, a distinctive attribute of the teachers with a propulsive conception of AT is that of an advanced re-conceptualiser. Such individuals will not be pleased with remaining at the status quo; nor will they be complacent about their students' good performance in the national examination. Instead, they will deal with the phenomenon by actively creating platforms for their students to go through a different teaching and assessment experience. They are often able to reflect on the current assessment practices and think beyond the current assessment culture and norms. Such teachers tend to think of the students' language learning beyond examinations, which is much long-term in focus.

To sum up this experience of AT, CL teachers experienced tension in assessment because of the multiple purposes it has to serve. They are experienced via (1) teaching accountability; (2) different roles of assessment; (3) their hope to impact students' future learning with assessment. These three different categories depict the different ways in which CL teachers understand AT in terms of assessment purposes. Accordingly, their different ways of dealing with the phenomenon are also described.

4.3.2 Conception of AT in terms of reality and ideal gap

CL teachers experience the tension of current assessment practice when it is compared with their ideal assessment practice. To this end, this conception may be understood in terms of assessment reality and ideal gap. The focus in this conception is on the various ways in which CL teachers experienced and understood the AT related to the current assessment practice and their ideal assessment. Similar to the previous discussed tension, three categories of description emerged, each depicting the variation

in which the teachers experienced the phenomenon, as well as a relationship between categories that delineates the increasingly complex ways of understanding this phenomenon, together with how they deal with these AT.

Category 1 (Passive): Teachers' assessment practices are restricted by the forms of current language testing

This category depicts how CL teachers feel passive and uncomfortable with the construct of the current language assessment practice, focusing particularly on the forms of language testing. These may refer to the weighting of the language skills (listening, speaking, reading and writing) tested in the SA, as well as how these skills are being tested. Here, while they focus on limitations such as the “discrete-point language testing” approach (Lado, 1961), they also think that some adjustments or changes could be made so that it will be more ideal and valid to assess their students' language learning. The following teacher expressed her sentiments as such below:

Extract 4:

*Teacher A: Language assessment, actually I felt that for each student, there are great differences among their language ability. Some are good at speaking. So if for me, I felt it is, **our national assessment in language is looking at listening, speaking, reading and writing holistically**. But I felt some, he could be one who is just good at speaking. So is there a possibility that his language ability is mainly exhibited in speaking? Is like there are many who studied Chinese but at the end not many turned out to be a writer.*

The teacher's focus here is solely on the construct of the national examination papers where the four language skills would be tested separately. Apparently, they feel that all these issues are not within their control. The teacher was also doubting the quality of this national test by questioning how valid it is to assess students' language ability using the “non-integrated” approach. Below is another example of teacher expressing the same concern:

Extract 5:

*Teacher E: For example, in Paper 1, I think there is no need to test a functional writing and a composition. **The two parts are testing the same skills.** Because one writing skill, by only test one component, we actually would be able to know where his skill is.*

The focus here is how many percentage or how many parts of the examination paper should be dedicated to assess students' language skills, such as writing. CL teachers' concerns in this category are still relatively narrow as their focus is on current limitation in the national examination, which focuses on students' short-term learning and achievement. Their focal awareness did not go beyond the validity of the "non-integrated" assessment approach in the national examination, which is beyond their sphere of control. In this sense, the teachers could only stay passive to just voice out their unhappiness and stay with the status quo with this arrangement in language testing. There are better ways to deal with the phenomenon, which will be addressed in the more sophisticated categories described in the subsequent sections.

Category 2 (Responsive): Teachers started to question the authenticity of the current language assessment approach and suggested to use the task-based assessment

This category is differentiated from Category 1 in that the conceptualisation of the AT by the teachers is now focused on the current opportunities to see how language assessment approach could be re-considered. Key attributes included the teachers' uneasiness over the "non-integrated" language assessment approach as they feel that the "integrative" or "communicative" approaches are more authentic and effective to assess or encourage their students' language learning. In addition, this category expands the focus to both national examination and classroom assessment, which still are on students' short-term learning and achievement. However, this is more complicated than the previous category as it requires a shift in awareness in language teaching and assessment approach. Teachers start to question the forms and functionality of the current language assessment, as well as the validity and authenticity of it. For example, some teachers are more responsive and could even suggest using task-based language assessment as their ideal assessment approach after discussing some of the

disadvantages of the current assessment practice, as exemplified by the teacher's response below:

Extract 6:

*Teacher D: The current assessment approach is on the listening, speaking, reading and writing. There can be variations in between but eventually still on these skills. If you really give me absolute freedom to think about my ideal assessment, I think language assessment should include argument as a mode. Because to argue, that is to let them see the difference in viewpoints and there will be 'thought collision', those are inner thinking which needs preparations. For example, I can give a current hot topic and divide the students into groups for them to prepare the information needed. They need to do research and present them into sentences and do presentation. They cannot any how say and thus the sentences must be refined. **So during this process, we can assess many abilities.** How do you cooperate? How do you do research? How to search on the Internet. And how to write it out?*

The teacher is more responsive than the previous category in that she sees the limitation of the current assessment practice but treats it as imminent opportunity and she goes on to suggest better ways to assess in terms of the function and form. In this case, giving students a task (an argument), which could activate their other living skills, thus making the assessment much more authentic and valid because it is believed that this assessment can assess many other non-cognitive abilities. Likewise, Teacher B suggests using task-based assessment because it can overcome the short-comings of the "non-integrated" assessment approach.

Extract 7:

*Teacher B: So for task-based assessment, I think it is **very close to students' lives and they really are using the language in an authentic setting.** So students need to learn to use. And use it in their*

*actual lives. I think the listening, speaking, reading and writing skills need to be tested. How do you test? They must be able to use, you must be **able to use it in an integrated manner and give them an authentic environment to activate their language ability to complete the task. That is the best evidence.***

The teacher with a responsive conception of AT tend to focus on his current assessment practices and their related challenges. The teacher viewed these as opportunities to change and improve his assessment practice. By asking questions on the functions and authenticity of current approaches in language assessment, he is reviewing the function of these current assessment practices. Subsequently, he suggested the “task-based language assessment” approach, which is more sophisticated but ideal (as it is more valid and authentic) to assess students learning, for both formative and summative purposes. He sees these tensions in assessment as some opportunities to change and influence the language assessment approach within his sphere of control. Unlike the previous category, the teacher in this category is looking at the imminent opportunity or gaps, which he is able to address. He is more interested in the experienced AT and would react to this phenomenon in an active manner. He adjusts his assessment practices. The teacher’s responses to the phenomenon seems to be guided by the intended outcome of his students’ language learning, which is also considered to be more short-term in focus when compared with Category 3.

Category 3 (Propulsive): Teachers begin to fundamentally re-conceptualise CL assessment beyond students’ current CL learning

This final category documents CL teachers’ experience and thus their ways of experiencing AT with the current traditional approach in language assessment that included both the “non-integrated” language assessment and task-based language assessment approach. This category has the most sophisticated focus on the conceptualisation of the tension than the previous two categories, which are situated at the current limitations and opportunities. Key attributes of this category include a focus on aspects which are beyond the current limitations and opportunities. Their concerns were not limited to only national examination and classroom assessment, but on a much

broader scope of assessment. In fact, their discussions went beyond the forms and functionality of language assessment. Unlike the previous category, this one went beyond validity and authenticity issues of assessment. To be more precise, this category captures the teachers' total overturning of what traditional or current CL assessment should assess. Their assessment focus is on students' long-term language learning. These may include non-language aspects of a students' language learning. This fundamental re-conceptualisation of CL assessment requires a total paradigm shift of what language assessment can and should assess, beyond what the schools are offering currently. In other words, teachers are talking about changing the beliefs and conceptions of assessment altogether. It is because of this propulsive force in the conceptualisation of assessment that allows CL teachers to fundamentally re-think about traditional language assessment and go beyond students' current CL learning and assessment, as illustrated by the teacher's response below:

Extract 8:

*Teacher E: I have this thinking because I found that when students learn, whether it is language learning or not, actually there is this 'inner quality' which, for example, this 'keep on fighting' spirit. Because of this quality, so this is also an influencing factor for him to learn the language. **So when you assess or help him to improve this inner quality so that he can transform it into language learning, I think this is it.** Students must have certain inner quality that allows them to fully capitalise on the assessment given by teachers. If not, this assessment would be useless.*

The teacher has re-conceptualised the traditional or current language assessment because her ideal assessment is to assess "other inner qualities" rather than just on language skills in her lessons. However, this inner quality (such as "keep on fighting spirit") would also have a direct impact on her students' language learning. Feasibility and applicability aside, she has gone beyond the forms and functions, as well as validity and authenticity of current language assessment. Below is another example:

Extract 9:

*Teacher G: Language teaching and assessment have to be **flexible and multi-faceted**. So I would hope this mini novel teaching can bring what kind of reading experience to the students, or let them associate with something in their lives. **You mobilise their emotions and then you allow them to write a paragraph**. They would be able to write. For this test, **it is beyond just language assessment**. And it is really practically assessing how the child's can allow his emotions to write into words, about how to express himself. This is very touching.*

Here, both teachers are deciding on the values of current assessment practice and approach. They are aware of the current assessment reality and have thought beyond students' current CL learning experience. In Singapore, and in many countries alike, language assessment has always been very standardised. Being "flexible and multi-faceted" is indeed innovation in language assessment which may make teachers struggle against the conventional force of the dominant norms and current practices. Using assessment to "mobilise students' emotion to write" in CL learning is a clear indication of a re-conceptualisation of current language assessment function and forms. The teachers are able to empower themselves to make professional decisions on assessment tasks and this propulsive force behind allows them to re-conceptualise the current language assessment practice. These are far beyond the previous category where the focus is only on the forms and functionality of language assessment, as well as the validity and authenticity of assessment. A distinctive trait of teachers with a propulsive conception of AT is that of a re-conceptualiser. Such teachers would deal with the gap between the current assessment practice and their ideal assessment with a strong belief. They are always ready to use this core value to create other possibilities that are often pleasantly surprising. AT drives and propels them to act with conviction when dealing with CL teaching and assessment. They always believe that there is something more they could offer in teaching CL, other than just guiding their students to perform well in the national examination. These teachers' assessment focus is on students' long-term learning, which is often beyond what the current schools can offer.

In summary, CL teachers experience of AT between the current assessment practice and their ideal assessment. While this tension starts off with a limiting or passive experience in Category 1, it goes on to the “responsive” level where teachers see the challenges in assessment as opportunities to make some changes to their practice. When teachers’ experience of AT eventually advances to Category 3, it has evolved to be a propulsive force which propels teachers to think deep and re-conceptualise their current assessment practice. Correspondingly, teachers assume the role of compliant executor, rational practitioner and re-conceptualiser when they come to deal with AT.

4.3.3 Conception of AT in terms of involvement of stakeholders of education

CL teachers experience tension in assessment when they feel that their professional decision making in assessment is being challenged and compromised by the involvement of stakeholders of education, which may include parents and students. Broadly speaking, they may also include schools and the community. CL teachers’ experience of AT with assessment demand from stakeholders of education created three categories of description. Similar to the previous two tensions, the categories progress in complexity as they advance. Here, the three categories demonstrate not only the progression from a narrower experience, but also demonstrate the hierarchical relationship between them.

Category 1 (Passive): Teachers remained passive and compromise their professional decision-making in assessment

Support, engagement and involvement of stakeholders of education are critical to the success and improvement of teaching, learning and assessment in schools. They can either affect the many decisions that take place in schools every day. While some of these demands are valid and sound, others may be detrimental and affect teachers’ professional decision-making. This category highlights the teachers’ concern over the current limitation of demand from students in assessment and their passiveness to this professional decision-making. Below is a good example of such a teacher’s experience:

Extract 10:

*Teacher B: I have these Sec 5 normal students. Oh, that is at the beginning, I told them the test marks are not important, what is important is the corrections. Then they didn't listen. The second time I told them again but they refused to listen again. And so the third time. Therefore, I felt that I was wasting their time, I was doing things which were thought of not important. Then, I didn't do it for every question. I still return their tests and of course in order to make my teaching more effective and **here if I cannot change the students, and they don't take your advice and think that it is not important, of course I have to make changes. So I will perhaps reduce descriptive feedback.** So from those important ones, for example those ABCD questions, I will just tell them the correct answers. But I used to explain and explain. So if they don't listen once or twice, then I will just provide the correct answers.*

Here, the teacher is affected by the students' responses to his assessment thinking and practice. He is making a professional decision in assessment but the students do not adhere to it. His students are more concerned about their test marks and are not interested in the corrections and descriptive feedback provided by the teacher. In the end, the teacher is controlled by this tension and fully compromised on the students request of not providing descriptive feedback. The teacher has conceded the situation and his assessment focus is on students' short-term learning. Consequently, the teacher stays passive and self-empowerment in assessment is not observed here. It can be concluded that the AT created by the demand of students limited the teacher's sound assessment practice. One of the clear attributes of the teachers who assume a passive conception of AT is that of a docile executor. Such individuals would not push for any changes in assessment, although he or she is fully aware of the benefits it can bring to their students' learning. They easily give up any sound practices in assessment when the stakeholders of education are sceptical about them. They are resigned to fate when it comes to coping with the demands from their stakeholders.

Category 2 (Responsive): Teachers partially compromise their professional decision making in assessment in response to the demand from parents

This category differs from Category 1 in that it encompasses more than just the current limitation (limited by students' assessment demand) for improving teachers' assessment practice and advances the AT focus to current opportunities, making it a more complex experience or conceptualisation of the tension. However, both Category 1 and 2's assessment focus is on students' short-term learning. The key attributes include the teachers' firm stand in their assessment. While the teachers may initially struggle with the demand from parents, they are responsive and view these as opportunities for them to reaffirm their current assessment practices. These teachers are more confident than the previous category when dealing with stakeholders' demand in assessment. They hardly make any compromise in their assessment practices, although they may sometimes meet with unreasonable demands from parents and students. They usually are able to stand firmly on their professional decisions, even if the situation is beyond their control, like the teacher below:

Extract 11:

Teacher C: And because of this assessment task, I had an argument with a parent. Because this parent reprimanded me for giving her child the "I love kitchen" project. But that student did not do. Then I told her the child did not do the project but I have to key in the CA marks. So no marks lah. At the end, the parent was not happy. That was rather complicated because there were other things and argument beforehand. Yes, then I told the parent there will be no mark and so on and I forgot the details because that was a while ago. At the end, she came to reprimand me during the parents-teachers' interaction day. Then, this parent is really, asked me why I gave her child this project task, she was very rude. She said "Why do you give this kind of assignment?" Then she placed her mobile phone on the desk, "what are you doing? The task you gave was so very tough. How do you expect my child to do?" Ha, I was thinking, this is considered tough? Many students had done that in the past. There is not much

requirement in the project, just so simple with some steps. I didn't expect you to cook or stir-fry or fry food. You can do a sandwich-making. She said I had given a tough job. What do you expect me to do? How to do? At the end, I didn't want to argue with her, so I got Michael (the level head) to help me to talk to her

The teacher had recounted and highlighted her unpleasant experience with a parent who reprimanded her for an assigned assessment task. While the parent demanded an easier assessment task, the teacher was not ready to compromise. Such interferences from stakeholders are not uncommon in Singapore schools. In reality, teachers often need to tactfully deal with these demands. Here, the teacher did not respond to the parent's reprimand directly but tactfully referred the case to a higher authority (the level head) in the school to seek further help. She had not compromised her assessment practice as she thought that it was professional and sound. She saw these "unreasonable" demands from parents as a force for her to continue her professional assessment practice. She continues to reaffirm herself to such assessment decisions and practice. She would like to continue with her professional decision in assessment and exert a force to put across her stand, even if it is beyond her ability and control.

In brief, for this category, CL teachers expressed their frustration and tension in assessment due to "unreasonable" demand from parents. While it might seem to be something negative, they are able to take this imminent opportunity to reassure themselves of the sound assessment practice which they have been strongly believed in. Unlike the previous one, this category progresses from the current limitations to opportunities as CL teachers seek to continue to believe in themselves in professional assessment decision-makings and practice. A unique characteristic of teachers with a responsive conception of AT is that of a rational practitioner. Such teachers possess the professional knowledge and confidence to deal with the phenomenon, even if the situation is beyond their control. They will seek help to put their sound assessment idea across.

Category 3 (Propulsive): Teachers reflect deeply and professionally about the role of parents in assessment beyond current school practice

This category is distinct from the previous category in that the conceptualisation of the AT is now centred at a much-sophisticated level. It is a paradigm shift of how involvement of parents in assessment could be taken upon. This AT was seen to be a propulsive force behind the teachers, whom are very eager to seek improvement in their assessment practice by involving parents. This category evidenced CL teachers' experience of AT as they re-conceptualise the parents' role in assessment, which somehow changed the conventional or traditional way of responding to the demand of parents. By doing so, the parents' involvement in assessment is now being conceptualised as a way to reflect and improve the teachers' assessment practice. Unlike previous categories, the teachers now believe that engaging parents in assessment is helpful. Below is a good example of such a teacher's experience:

Extract 12:

Teacher I: So if, I was thinking, I was thinking, if our teaching and assessment can involve parents to give feedback, I think it will be rather good. But this idea will make many teachers 'jump'. If assessment can involve parents, that is come in to play a supervision role, I think it will be better. But teachers may not..... Sometimes, if we let teachers, for example, if they are bolder enough, that to say, ah, "Dear parents, I am the CL teacher of 1E1, if there is any gap in my teaching, do inform me directly". Ah, then if parents have any feedback, that is also part of assessment. I am not sure if teachers can accept this? Interactive. Ah, I did try. I would inform parents that for this term, I have given ten assignments and so on. Your child has only handed in three assignments. Could you help me with the rest?
I did this, so if the parents were able to help me take note of their children's learning, then they will be more serious.

The teacher has shown her professional and advanced thinking about the parents' role in her assessment practice. This re-conceptualisation of the role of stakeholder (parents) in assessment is a positive move. Instead of focusing on the current limitations of stakeholders' demand in assessment, she challenges this conventional paradigm and re-

conceptualised the role of parents in assessment. Teachers with propulsive conception of AT will readily welcome and attempt any alternative ways of improving their assessment practices and subsequently their students' learning by strategically involving the stakeholders of education, often beyond schools' capacity. As their ideas are often "out-of-the-box", other teachers may not be ready to accept their ways of dealing with the phenomenon. However, they are daring enough to suggest "re-conceptualised" ideas that are radically different from the conventional ways of doing assessment in schools. One of the prominent attributes of teachers who have a propulsive conception of AT is that of an advanced re-conceptualizer. Such teachers will tend to look at the long-term learning effect of their assessment practices when dealing with the tension in assessment. They usually have a strong belief in themselves and think that they can always make things better, both for the teachers and their students.

In short, this AT demonstrates the different ways in which CL teachers experience tension with stakeholders of education. This variation in experiencing AT is differentiated by their focal awareness and how they expressed their experience. While teachers with a passive conception of AT focus on what they cannot do in terms of professional decision making and teacher empowerment, the teachers with a responsive conception of AT focus on the opportunity to exercise their professional judgment and empowerment in assessment practices. However, the propulsive conception of AT is the most sophisticated because it goes beyond teachers' professional decision-makings and empowerment. Such teachers are strong believers who are able to re-conceptualize assessment issues to improve current practices for the long run.

4.4 Summary

In summary, this chapter has reported three distinct but related ways of teachers' experience of AT. The three conceptions of experiencing AT are the passive, the responsive and the propulsive.

It was found that the passive conception of AT denotes that the teachers' focus is on the current limitations and they are restricted by many forces, which are not within their sphere of control. Such teachers usually stay with the status quo and are not willing to change any assessment practices. The responsive conception of AT is more

sophisticated as the teachers' focus is on the imminent opportunities. Such teachers view tensions in assessment as opportunities for them to take up ownership to make changes in practice. Both the passive and responsive conception of AT focus on students' short-term learning in school. The propulsive conception of AT is the most complicated in the outcome space as the teachers' focus is beyond the current limitations and imminent opportunities of their assessment practices. They are more concerned about students' future and long-term learning. Such teachers often look forward, reflect and re-conceptualise assessment issues beyond schools. AT is understood to be a constructive and useful force. While Conception A and B's assessment focus is on students' short-term learning, Conception C's focus is on students' long-term learning.

These three conceptions of AT form the outcome space of the research in that the degree of sophistication increases from Conception A to C.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Recapitulation of research purpose and findings

In the previous chapter, I described the qualitatively different ways in which nine CL teachers with various profiles had experienced AT. These discoveries were presented as three conceptions, which are theoretically and logically related to each other in a hierarchical fashion. Collectively, the conceptions make up the “outcome space” that represents the full variations of the themes and critical aspects (the ‘What’ and ‘How’ aspect) experienced by the teachers in AT. This is the first research known to investigate conception of AT held by CL teachers in Singapore and is valuable because school leaders and policy-makers can be better informed of the current AT experienced by teachers.

To recapitulate, this research investigates the conceptions of AT of Singapore CL teachers. Drawing on the literature of AT, teachers’ conceptions of assessment and teachers’ experience of assessment in Singapore, a case is made to explore in greater depth about teachers’ distinct experience of AT using phenomenography. CL teachers were chosen as participants due to their varied backgrounds that include their nationalities, academic and teaching trainings. It is argued that these teachers’ conceptions of AT and how they deal or manage tension in assessment is of critical importance in the future assessment policy planning, initiatives, reviews or reforms, as well as teacher trainings and practices.

Using semi-structured interviews, interviewees were led by me with pre-designed interview questions and sub-questions to reiterate and to reflect on their experience of AT. Following the conventional method of phenomenographic analysis, the teachers’ interview data were analysed using iterative and inductive processes. At the end, the three conceptions of experiencing AT are found to be the passive, the responsive and the propulsive. Correspondingly, they assumed the role of the compliant executor, the rational practitioner and the advanced re-conceptualiser when dealing with AT in their practices. The entire complexity of these conceptions encompassing all the different ways of experiencing this phenomenon formed the “outcome space”.

5.2 Significance and contributions of the research

In this section, I discuss how the findings of the research, including the ‘What’ and ‘How’ aspect of teachers’ experience of AT, compare with past studies. More importantly, I highlight how the more sophisticated conceptions of AT make contributions to the new knowledge of assessment practice and experience.

This research argues that AT should not always be assumed to be problematic and undesirable. In addition, teachers show a variation in their conceptions of AT. It depends on how teachers experience AT and how they deal with it. These findings allow policy-makers and school leaders to have greater awareness of how changes in assessment and the multiple purposes of assessment that can bring about much tension and teachers’ varied experiences are valuable to us. Fullan (2001) reminded that each individual experiences change in a different way. As such, teachers’ collective experience in AT can allow policy-makers and school leaders to make informed assessment about the antecedent conditions that are needed for teachers to ‘better’ face changes and challenges in assessment. In addition, I also argue that teachers need to be more aware of what can be expected of them in experiencing AT, if they are to adapt their assessment strategies and practices to the current demanding teaching landscape and improve their experience in assessment. To this end, the findings in this research provide a framework that can be used to further discuss and improve AT issues. It may also serve as a framework to account for how and why some teachers can have ‘better’ AT experience and ‘better’ ways of managing AT that was related to the kinds of pattern of variation that were being experienced.

Precisely, these findings are useful to offer valuable insights into the following four areas of educational assessment literature.

5.2.1 Understanding AT to be positive and constructive

The main finding in Conception B and C in this research directs me to rebut the presumption that AT is negative and thus I argue that there should be a new understanding of AT beyond this negativity. AT could be positive and constructive to teaching.

As reviewed in Chapter 2, there is a general consensus in the field that tension in assessment is negative and thus undesirable to teaching. For example, Bonner’s (2016) review on teachers’ perceptions about assessment concluded that when AfL and

accountability are in tension, the internal values may be suppressed, teachers will be unable to adopt good assessment strategies and thus get frustrated about the disconnection between their internal values and system values. This may eventually lead to teacher burnout and them quitting of the profession. Similarly, Bishop et al. (1999) and Lindsay and Lewis (2003) reported the situations of AT in the United Kingdom when primary or secondary teachers were asked to use assessment for different purposes. Bishop et al. (1999) studied users' perception of the GCSE and found that teachers experienced tension with the formative objectives (for improvement) of the GCSE while it was summative in nature. Teachers found difficulty in reconciling the fundamental approaches of balance, challenge and relevance for all within the GCSE with outcomes of public accountability (Bishop et al., 1999, p.48).

These cases resemble Conception A in this research which depicts the conceptualisation of tension in assessment as current limitations, which chained teachers to stay passive and do only what is within their immediate control in assessment practice. This tension confines teachers to have no desire to change the current status quo in their assessment practice, favouring SA over FA. This may be due to the "exam-oriented" teaching practice that is deeply rooted in the CHCs. In this conception, AT is experienced as an inertial and restrictive force. This conception shares similarity with many other previous AT research findings. For example, Lindsay and Lewis (2003), considering the multiple purposes of the Baseline Assessment Schemes, which is a statutory requirement for all primary schools in England from 1998, discussed the tension between the "pedagogic" and "managerial" purposes of the assessment that limited teachers from carrying out effective assessment and teaching because of the contradictions involved. For such scenarios, Black (2003) argues that unless the system of SA for accountability and certification of students is fundamentally reformed, the great potential of classroom FA for improvement in standard cannot be fully realised. His statement hints of tension for teachers with the current assessment system favouring public accountability. In particular, Conception A also echoes Carless's (2012) AfL research which pronounced that there is AT between helping student learn and grading for certification.

Specifically, AT has been seen to be undesirable as it may undermine teacher empowerment. Teachers may face the problem of inconsistency and even compromise

when making assessment decisions. More importantly, the variation in teachers' experience of AT in this research also rebuts this past presumption of AT being negative and undesirable. Conception B and C in the outcome space shed new light to a deeper understanding of the complex relationships among different assessments. They have shown that tension in assessment could be a positive and constructive force for teachers to reflect and change their current assessment practice.

While past studies have reported teachers' experience of AT due to many different reasons, this research found that teachers could still manage their assessment well despite their experience of AT.

Conception B and C are more sophisticated than Conception A as they surpass the focus of current limitation and conceptualise AT as opportunities to make changes to their assessment practice. The key attributes of these conceptions include the perception of AT as a force for them to take ownership and rethink about their assessment practices. It enables teachers to face the challenges in assessment by making some changes in their practices and even refocus students' learning and assessment in the longer run. This finding is particularly helpful since tension in assessment has been and would continue to be a common and prevalent phenomenon that is inevitable for teachers. It is thus more useful and productive to look at how teachers could 'better' experience AT, instead of trying to avoid or reduce it. As Dewey (1904) has identified the relationship between theory and practice in education, the tensions between the two are unavoidable but also a positive force for teachers' learning. Baumfield (2016) further added that if such relationship of theory and practice is "handled in such a way that is mutual and fortifying, then the fact that tension is inescapable need not be a problem and attempts to resolve it by privileging one or the other are misguided" (p.159). Such discussions on tension as a desirable force for teachers share similarities in Conception B and C where teachers did not seem to have favoured one or the other in assessment. In these conceptions, not only did tension not appear to be a problem, it looks more like a motivational force moving and propelling teachers to advance themselves in assessment practice.

Overall, the variation of teachers' experience of AT inform that the previous view about AT being negative may be too simplistic and stereotyping as teachers with greater awareness in their ways of experiencing AT show increasingly more useful ways of

managing tension in assessment. These more advanced ways of experiencing AT can be used to predict how new assessment ideas may be taken up by Singapore teachers. For example, in contrast to Carless's acknowledgment of the limitations of formative use of summative test (FUST) and the tension faced by teachers in the examination-oriented CHCs, this research provides meaningful insights into teachers' experience of AT bought about by new assessment ideas and practices. In my context, they were well-received by the "rational practitioner" in Conception B and "advanced re-conceptualiser" in Conception C respectively because AT has been useful and constructive in their experiences.

5.2.2 Increasing importance of FA in a SA-dominated educational landscape

From teachers' different ways of experiencing AT, I argue that there are varying degrees of importance of how FA is being conceptualised and practised in an educational landscape that is SA-dominated. Over the years, FA and SA have been seen to be a "binary" and in tension because of the specific purposes they serve and different "point in time" they are being carried out. Scriven (1967), Bloom (1970), Rowntree (1987), Broadfoot (1996), Boud (2000), Harlen (2006), Newton (2007), Moss and Brookhart (2009), Carless (2011) and many other scholars have discussed the aims, uses, functions, roles or purposes of assessment from a variety of perspectives.

Carless (2009), citing Boud's (2000) idea that an assessment has to do "double duty", explains that an assessment may need to serve grading (summative), as well as to enhance learning (formative) and therefore it is "beset with tensions and compromises" (p.79). Boud (2000) outlines some of the ways in which assessment has to perform "double duty" that include FA for learning and SA for certification. So there is the issue of whether an assessment is fit for its purposes. Naturally, the information gathered that are fit to be used for one particular purpose may not be suitable to be used for another. Even when there is a need to do "double duty", the purposes may still clash in terms of its impact and consequences (Mansell, James & the Assessment Reform Group, 2009). As the same assessment is performing another duty, teachers would need to consider the prioritisation issues in terms of the purpose (formative or summative), focus (immediate task or lifelong learning) and management (learning process or content domain). These multiple considerations naturally create tension for teachers.

These experiences of AT informed that there is a need to reconsider assessment being fit to serve a single purpose only.

My findings confirm these past studies on the tension between formative and summative purposes of assessment. Firstly, some CL teachers see FA and SA as two distinctive and separate entities which perform different functions. Similar phenomenon was observed earlier by Taras (2009). She confirmed that SA and FA have developed into two contradicting purposes in many instances. As she explains: “Since Scriven drew the distinction between SA and FA, a gradual separation of the two into mutually exclusive entities based on the differences in functions of assessment has evolved in the literature” (p.59). Such experience was identified in Conception A where teachers are limited to practicing only SA but not FA because of accountability. The figure below shows the focal awareness of teachers from Conception A.

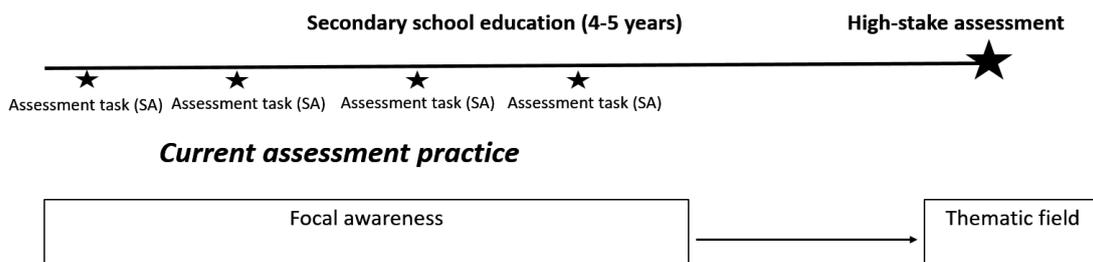


Figure 5.2.2a: Focal awareness and thematic field of teachers in Conception A

AT evolved when there is a call for assessment change to teachers’ current practice that mainly consists of SA. This experience of AT (the focal awareness) is directly related to the accountability issue embedded in the high-stake national examination (the thematic field). Teachers can only practice SA and FA has no place in this conception. Conception A shows that solely paying attention to the different purposes of assessment has brought about unhelpful tension.

This narrow and limited view on the purpose of assessment is not uncommon in practice, both in Singapore and elsewhere alike.

SA, usually represented by standardised and high-stake examinations, is often seen to carry certain negative connotation because of its undesirable backwash effect. Some studies (Reay & Wiliam, 1999; Sharpe & Gopinathan, 2002; Triggs & Pollard, 2000)

pointed out that many examinations directed teachers to “teach to the test”, paying special attention to what is going to be tested and structuring assessment during lessons to train students for the national examinations. As a result, teachers seldom use these assessments formatively. These practices may be further reinforced by the “good” results shown in the examinations. Evidence provided by Gordon and Rees (1997) showed that teachers have their ways to train students to perform well in any examinations, including those which assess higher-order cognition and skills. These cases echoed teachers’ experience in Conception A where they expressed that when it comes to examination preparations, they will ensure that students know the scope and requirement of these high-stake examinations. They will give plenty of assessment practices which are designed according to the national examinations. These teachers seldom do anything beyond examination preparations, such as those task-based language assessments. For such a phenomenon, Black and Wiliam (1998c) have commented that these SA would not accurately assess students’ knowledge level because students are able to answer these questions with memorization and drilling. Such teaching and assessment encourage superficial learning and would hardly promote higher-order thinking. Indeed, Pellegrino, Chudowsky and Glaser (2001) also questioned if SA could accurately assess students’ complex knowledge and skill acquired.

In all, SA is viewed as a powerful force which is capable of directing and limiting teachers’ everyday teaching in the classrooms, as shown in Conception A. Indeed, past studies tend to perceive SA and FA to be strongly at odds (Black & Wiliam, 1998; Holt, 2005; Sadler, 1989, Scriven, 1967; Stobart, 2008).

As reviewed in Chapter 2, advocating a more “balanced” assessment practice or system has been popular in recent years, both in Singapore and elsewhere. However, this notion has been hardly possible to achieve for teachers as they are unable to check if any assessment practice is “balanced” against some “standard” (if there is any). Indeed, since the idea was introduced, there has been hardly any report or research that evaluates the taking up of “balanced assessment”. One of the reasons is that this idealistic and condensed concept has been avoiding the complex realities of assessment having to serve different purposes. This notion is too simplistic because there would not be any concrete criteria and standard to show that any assessment practice is “balanced”. Thus, it becomes highly subjective for teachers to operationalise this notion.

They may attempt to “balance” their assessment between FA and SA; they may also focus on adjusting their assessment practice between quantitative and qualitative. Further, they may also shift their assessment from “standardised” ones to more “personalised” ones; or change from “teacher-centred” to “student-centred”. I have argued in Chapter 2 that these assessment complexities are inevitable and should not be approached by such a simplistic notion. While many others may not have used the concept of “balanced assessment”, they have suggested to better integrate SA and FA (such as formative use of SA and vice versa). Nevertheless, the Graded Assessment Schemes (GAS) in England (Noss, Goldstein & Hoyles, 1989) and the School-based Assessment in the Hong Kong Certificates of Education Examination (Carless, 2011) did not seem to suggest that integrating SA and FA is an easy task. Indeed, the binary and contrasting paradigms between SA and FA, according to Moss, Girard and Hanniford (2006), is not simply an issue of considering different purposes, methods or ways of assessing students’ performance; it is philosophical.

While I think there is a need to recognise the differences in aims, purposes and characteristics between SA and FA, it is more crucial to find a way of relating them together that preserves their different purposes while teachers are experiencing AT between these two purposes of assessment. This research argues that it is not necessary, and indeed it is not helpful, to be concerned with strict definitions and functions of SA and FA, as well as to “balance” them. There is good evidence from Conception B to show that teachers experienced AT in a ‘better’ way when their focal awareness shifted from the binary purpose of assessment (SA and FA) to how FA can enhance students’ learning in an SA-dominated educational landscape. Particularly, teachers’ experience of AT (of how they could combine SA and FA) has allowed them to consider using both assessments to work together towards productive teaching and learning. By allowing FA to play a more important role in their teaching, teachers no longer solely see SA and FA as “opposite” or serving “conflicting” purposes. This conception shares much similarity with the notion of learning-oriented assessment (LOA) which “seeks to contribute to the reconciliation of formative and summative assessment tensions by focusing on good assessment principles potentially applicable to both” (Carless, 2009, p.83).

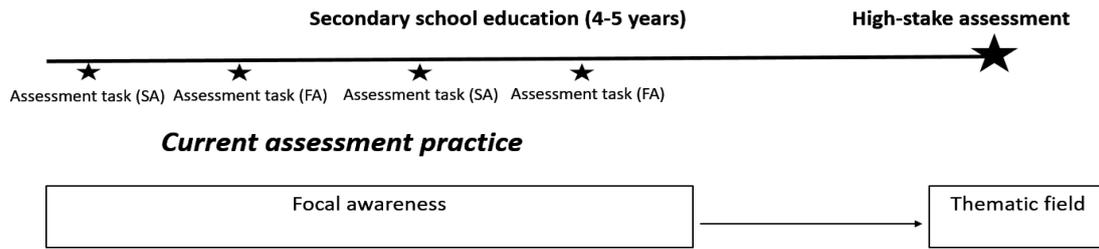


Figure 5.2.2b: Focal awareness and thematic field of teachers in Conception B

Teachers’ focal awareness in Conception B is on their current assessment practice, which includes FA and SA, with the national high-stake examination as the end point. In contrast to Conception A, teachers’ conception of AT shifted from a “zero-sum-game” between the purpose of FA and SA to a “complementary” relationship between the two in Conception B. There is an increasingly important role that FA could play in Conception B. The level of ambition for assessment practice from Conception A to B increases as teachers become more aware of the importance of FA for learning, with the national examination as the background reference point embedded in the thematic field. Both Conception A and B’s assessment focus is still on students’ short-term learning and achievement. This is clearly articulated by teachers in Conception B when they mentioned that they must overcome the challenges of using assessment. Using FA has made their teaching become more “fantastic”. They shared that FA has allowed their teaching to be more effective and at the end they also believe that students’ national examination results will not be poor.

Teachers’ experience in Conception B resonates with many other past studies, which concluded that FA does help in the performance of SA (Andrade, Du & Wang, 2008; Biggs, 1998; McDonald & Boud, 2003; Wiliam, Lee, Harrison & Black, 2003). Given the contrasting teachers’ experience of AT from Conception A to B, it seems to suggest that the ways in which teachers experience and deal with AT determines the possibility and the extent of the actual use of FA and SA. Teachers are able to include greater awareness of more aspects of experiencing assessment as it advances from Conception A. To be more precise, teachers are more aware of the importance of FA and could also see the relationship with SA. As such, they have ‘better’ experience of AT. Teachers’ experience of AT in Conception B resembles those which were reported in the United Kingdom.

Webb and Jones (2009), employing activity theory as the analytical framework, explore the tensions in developing assessment for learning in the UK. They have identified contradiction and tension in the activity system between teachers' beliefs about teaching and learning and current classroom culture and practice in assessment. In addition, they found that these tensions and challenges provided driving forces for cultural change in teachers' assessment practice. From teachers' experience of AT, my research goes further to discover the varied degree of evolvment of FA against SA in an "exam-oriented" teaching culture. It also found that some teachers could even project and visualise students' learning beyond the current stage of assessment, going into long-term, lifelong learning and sustainable assessment in Conception C. This will be discussed in the next section.

5.2.3 Re-looking AT in sustainable assessment

Boud (2000) defined sustainable assessment as assessment "that meets the needs of the present and [also] prepares students to meet their own future learning needs" (p.151). It is also known as the "third purpose of assessment"—assessment to foster learning throughout life (Boud & Falchikov, 2006, p.400). This assessment has to do "double-duty" because while it requires to first serve the present assessment needs of formative and summative purposes, it also needs to prepare students for their future learning (Boud, 2000). To this end, teachers experience much AT as sustainable assessment needs to accommodate both the present and future needs of students, which may not always be compatible to the current context and practice.

In general, this research found that teachers in the more advanced conceptions of AT are more ambitious in using all kinds of assessment to prepare students for lifelong learning because their focal awareness is on students' long-term learning and achievement. While Conception A and B is still very much concentrating on assessment in schools, Conception C has advanced beyond students' learning in school. A more advanced way of experiencing a phenomenon, according to Marton and Booth (1997, p. 107), means that "more complex and more inclusive (or more specific) than less advanced ways of experiencing the same thing". This puts me in a better position to reconsider the notion of "double duty" in which both FA and SA are not in harmony. Conception C suggests that when teachers have more awareness of the aspects of

experiencing AT, they could better see the complementary relationship between SA and FA. This finding supports Leong and Tan’s (2014) conclusion that the institutional authority of successful high-stake examination results is being challenged under the new policy that promotes FA.

While FA is mainly to help students improve their learning, SA is usually to account and certify how well students had learned. FA is commonly known to address the immediate needs of students’ current learning (such as the use of feedback) and sustainable assessment is believed to be able to contribute in some way to their prospective learning (Boud & Falchikov, 2006).

Over the years, SA and FA have created much tension for teachers in practice and sustainable assessment is a way of thinking to refocus these assessment practices. The findings suggest that the ways in which teachers experience AT will determine how well sustainable assessment will be taken up.

As Boud and Soler (2016) have recently reviewed, most of the past studies endorsed or used the idea of sustainable assessment to discuss other considerations of teaching and assessment. However, they also noted that this notion has yet to have a widespread impact on assessment discussions and the potential of sustainable assessment is still to be fully realised. Boud (2000) suggested that this assessment must do “double-duty” because while it requires to first serve the present assessment needs of formative and summative purposes, it also needs to prepare students for their future learning. To this end, there is a prioritization issue of putting FA and SA first before sustainable assessment could work in harmony with them. It is important to note that teachers’ experiences of AT in Conception C show otherwise because a focal awareness that includes SA, FA, national high-stake examination and students’ future career-related assessment is discovered.

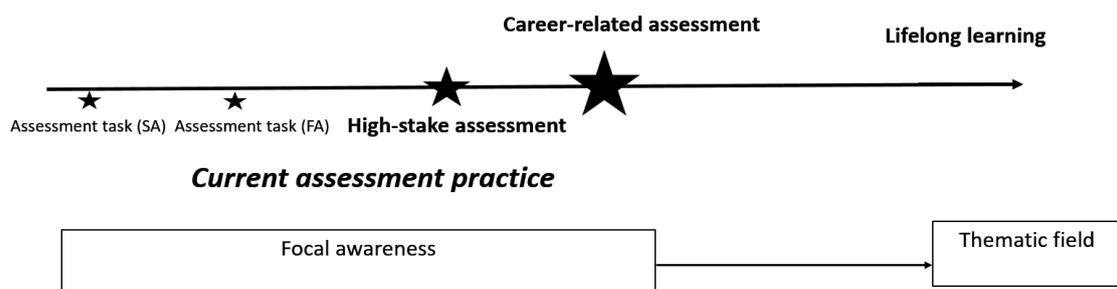


Figure 5.2.3: Focal awareness and thematic field of teachers in Conception C

Unlike Conception B, the end point (if there is any) for teachers' point of reference embedded in the thematic field in Conception C is not the national high-stake examination but their students' future (lifelong) learning. In contrast to Conception A and B, Conception C is the most sophisticated in the outcome space as teachers' focal awareness is beyond current limitations and imminent opportunities. Instead, it is on students' future career-related assessment. Teachers feel challenged when they must use assessment to prepare their students not only for the national examination, but also for life. They exhibit a paradigm shift in what they traditionally believe classroom teaching and assessment should entail. Conception C focuses on students' future use of the learned language, on top of doing well in the national examination. When juxtaposed with Boud's (2000) notion of sustainable assessment, the similarities between his idea and Conception C are very striking.

	Sustainable assessment	Conception C of AT (in terms of assessment purposes)
Main idea/focus	<ul style="list-style-type: none"> • Assessment 'that meets the needs of the present and [also] prepares students to meet their own future learning needs' • Double-duty 	<ul style="list-style-type: none"> • Teachers experience AT but could still focus beyond current limitations and imminent opportunities • Focus on students' future use of the learned language, besides doing well in the national examination
Relationships with FA and SA	<ul style="list-style-type: none"> • Re-focus FA practices • Balance FA with SA 	<ul style="list-style-type: none"> • Leverage on all kinds of assessment to improve students' learning • Let their students taste success (using FA) while preparing them for national examinations (SA)
Moving on	<ul style="list-style-type: none"> • A repositioning of assessment as an integral 	<ul style="list-style-type: none"> • Re-conceptualise assessment and actively creating platforms for students to go through a

	part of curriculum and pedagogy	different learning and assessment experience
--	---------------------------------	--

Table 5.2.3a: Comparison between sustainable assessment and Conception C

As reported in last chapter, teachers from Conception C shared that although their present goal of using FA is to help students do well in the national examination, their ultimate big goal is to see them use CL as a career or profession, such as a DJ in a radio station. They hope to use assessment to help their students to be among the best in CL and could use the language as a tool for their careers in the future. This account shares many similarities with sustainable assessment, which plays the crucial role of ensuring that teachers do not only meet students’ current or short-term needs, but also for their life-long or future learning.

In addition, Conception C informs that teachers’ experience of AT could help them re-focus their assessment practice towards sustainable assessment. Table 5.2.3a shows that in terms of the idea and focus, both are very forward-looking and require assessment to do “double- or multiple-duty”. Their relationships with FA and SA are also concrete for they serve different purposes to improve and support students’ long-term learning. As they move forward, they are reflective and would make changes to position assessment to support students’ lifelong learning that includes career-related assessment.

Boud (2000) has made recommendations for teachers to take on the sustainable assessment agenda and for further building its framework. His suggestions include taking on a criterion- or standard-based approaches, to have the belief that all students can succeed, students must also believe that they can be successful, to separate feedback from grading, to emphasise students’ learning rather than performance, to develop self- and peer-assessment and lastly to close the feedback loop. These suggestions imply that the current teaching and learning practices need to be changed. However, changing teachers’ and students’ current practices require a re-focus in assessment thinking which brought about much tension to these practitioners. For examples, for teachers to develop students’ self-assessment ability, there is a need to fundamentally change their current teaching practice. In the same vein, for teachers to

believe that all students can succeed, there must be a shift in their conception about teaching and learning. Not only do these recommendations require time and effort, they will also require teachers to change their current practice and thinking about assessment. This inevitably invites tension.

Recommended practices to support sustainable assessment (Boud, 2000)	How recommended practices are in tension?	How do the findings of teachers' experience of AT illuminate or inform sustainable assessment practice?
A criterion- or standards-based framework is necessary	A change in assessment practice invites tension	<ul style="list-style-type: none"> • Teachers' experience of AT may not always be negative and undesirable. • Good evidence to show that there is no need to reconcile the tension between FA and SA before sustainable assessment could be achieved. • Assessment needs not do "double-duty" to support students' lifelong learning.
The separation of feedback from grading		
The development of self-assessment is vital		
To complete the feedback loop		
Reflective assessment with peers should be encouraged	A paradigm shift in assessment creates tension	
The focus of assessment should be on learning rather than performance		
A belief that all students can succeed is needed		

Table 5.2.3b: Sustainable assessment practices in tension

Exploring teachers' experience of AT becomes more meaningful and productive since Boud (2000) explains that it will be challenging for teachers to take on sustainable assessment, both at the practical and philosophical level. This is because changing teachers' current practice and thinking about assessment creates much tension. Indeed, as described in last chapter, teachers experienced the tension brought about by the formative and summative purposes of assessment in different ways. However, I argue that these insights of teachers' experience of AT could directly inform scholars and policy-makers of how sustainable assessment has been taken up in Singapore at the pre-tertiary level and how these AT experiences would help or hinder the development of sustainable assessment.

Firstly, Conception B and Conception C show that tension in assessment could be constructive and even desirable when teachers are able to manage and respond to it in a positive fashion. Secondly, there is good evidence from Conception C to show that there may not be a need to reconcile the tension between FA and SA before sustainable assessment can take place. For example, teachers' (in Conception C) experience of AT has been constructive because it allows them to reflect on the purposes and impacts of their teaching and assessment practices to the students in the long-term. These teachers do not privilege SA or FA, but emphasise more of their students' lifelong or future learning. They are very ambitious in using all types of assessment to enhance their language teaching so that students could use CL with an acceptable proficiency in their future career undertakings. It suggests that it is not necessary to first balance SA and FA, as well as to handle the "double-duty" issue before sustainable assessment could happen. To this end, policy-makers should be more optimistic in advocating sustainable assessment at the policy and school level. Indeed, the recent announcement by the Ministry of Education (MOE) on 'Learn For Life' – Preparing Our Students To Excel Beyond Exam Results, too suggests that Singapore students need to become lifelong learners to meet the demands of the increasingly complex world ahead. In order to build the foundation work for nurturing life-long learners, MOE is making encouraging move to reduce school-based assessment load (MOE, 2018). To this end, the findings in this research are highly relevant to the current education context where AT exists between FA and SA but sustainable assessment was very much undermined.

5.2.4 Re-discovering the open tradition of Confucian education from CL teachers' experience of AT

As reviewed in Chapter 2, the taking up of new assessment ideas (such as FA, AfL, sustainable assessment and so on) in the CHCs have been challenging and are filled with tensions (Carless, 2012; Davison & Hamp-Lyons, 2010; Deng & Carless, 2010; Hamp-Lyons & Davison 2010; Harlen & Deakin Crick, 2002; Luk-Fong & Brennan, 2010; Thanh-Pham & Renshaw, 2014). However, there is a dearth of study on how teachers in the CHCs experience and manage these tensions. The current research sheds light on how Singapore CL teachers experienced AT and their ways of managing the tensions. To a large extent, it adds to the corpus of work on how new concepts in assessment are being taken up in Singapore, a CHC educational jurisdiction which shares many similarities in culture with other CHCs and yet has its distinctiveness. To this end, CL teachers are the best representatives among all subject teachers to characterise how a CHC would respond to new assessment ideas from the 'West'.

Contrary to many popular beliefs that Confucian education is rigid and not readily open to other ways of teaching and assessment, Tan (2017) found that Confucian education actually engages with other traditions, learns from all sources and adapts to changing times. She termed this tradition as open tradition in Confucian. Drawing evidence heavily from the *Analects (Lunyu)* and *Xueji (Record of Learning)*, Tan (2017) concluded that:

An open tradition ensures that Confucian education is not essentialised, static and fossilised. Instead, it is diverse, fluid and evolving, offering an educational paradigm that is all-rounded, ethical, universal and ultimately enduring (p.11)

To a large extent, CL teachers' varied experiences of AT provided another source of evidence for readers to re-discover this "open tradition" of Confucian education, which previous scholars did not see such cross-cultural exchanges and interactions to be aligned with the CHCs. In particular, Conception A confirms major past AT studies done in the CHCs (mostly in China) where they reported different degree of tension and difficulty teachers experienced when they try to incorporate FA (or AfL) into their

teaching. It becomes a barrier to their teaching. Teachers in Conception A view using FA and SA as “zero-sum game” in which FA's gain is equivalent to SA's loss. Previously, Chen et al. (2014), Gu (2014), Liu and Feng (2015), O'Dwyer (2016), Yan (2012), Yin and Buck (2015), Xiao (2017) concluded that the cross-cultural borrowing of FA from the ‘West’ is problematic in China because of the differences between “Chinese” and “Western” notions of knowledge transmission.

Turner (2011) explains that due to their Confucian inheritance, students “value collective harmony, refrain from expressing opinions that assert their own individuality in thought but that might clash with group norms, and habitually defer to authority” (p. 105–106). As such, CHCs usually favour teacher-centred pedagogy, and place an emphasis on academic performance in high-stakes examinations, with teachers taking big responsibility of students’ academic results (Luk-Fong & Brennan, 2010). Indeed, teachers in the CHCs are always assumed and expected to help their students do well in terms of good grades rather than on their students’ learning holistically. The conceptual shift in the different roles of assessment from summative to formative purpose, as well as those new initiatives such as “balanced assessment”, have been delivering tension to teachers. This could be observed from the assessment experiences of the nine teachers who participated in this research, albeit with variation in the ways they experienced AT. Poole (2016) thinks that some of the barriers include student and teacher resistance to these new practices of FA and the notions of “face” (面子 or dignity). These challenges are found to be directedly related to the deeply-held historical and cultural beliefs of teaching and learning that favour knowledge transmission and respect for the teachers. In the same vein, Carless and Lam (2014) also believed that the 20 centuries-old history of competitive exam-oriented culture needs to be taken into careful consideration in the discussion of contemporary assessment practices in Confucian-influenced societies.

However, teachers’ experiences of AT from Conception B and C also confirmed that there is a limit to how the above-discussed historical and cultural perspective can be used to explain the taking up of new assessment ideas from the ‘West’ in the CHCs. Instead of concluding that the ‘Western’ ideas of assessment are incompatible and in conflict with the CHCs, teachers’ experiences of AT in the more advanced and sophisticated conceptions show that these ideas and practices need not to be incompatible in the infamous exam-oriented CHCs. As discussed in last section,

Conception C is so sophisticated in that it is neither necessary for teachers to balance SA and FA, nor to handle the “double-duty” issue before sustainable assessment could take place. In other words, there is already good evidence from teachers’ ways of managing AT to show that although there are tensions and challenges in taking up new assessment ideas from the ‘West’, teachers in Singapore have shown to be very open and ambitious in using these assessments in their teaching. This is the “open tradition” of Confucian education that is largely unheard of from many past studies.

To this end, Conception B and C also seem to be echoing Goulah’s (2010, p.42) notion of the “East-West ecology of education” and He’s (2013, p.62) idea of the “East~West epistemological convergence of humanism” as they believed that there are common grounds and spaces for both ‘Eastern’ and ‘Western’ ideology of education to interact and even encourage intercultural dialogue with different thinking in educational practices. In fact, Goulah (2010) seeks to rediscover a new humanism (that includes language, culture and education) common to both the ‘West’ (represented by Lucy S. Mitchell, Francis W. Parker and John Dewey) and ‘East’ (represented by Tsunesaburo Makiguchi and Daisaku Ikeda). Similarly, He (2013), drawing on the works of Confucius, Makiguchi, and Dewey, considers “East~West epistemological convergence of humanism” as:

*an epistemological foundation for theorizing languages, cultures,
and identities in education to fostering fruitful interactions that
commit to a high level of human potential and exhilarate
intercultural dialogue with difference (He, 2013, p.62)*

The above recent studies on the East and West thinking in education suggest less of a strict dichotomy between the two cultures. Instead, there is much more room for both cultures to interact and even to draw on each other's strength for further development. My research confirms such a phenomenon. Teachers in Conception B and C deeply believed in the importance of national examinations to their students’ education, a typical trait of the CHCs of the ‘East’. However, their experiences and management of AT also revealed that they can integrate assessment concepts (such as FA, balanced-assessment and sustainable assessment) from the ‘West’ into their exam-oriented

teaching. This “open tradition” of Confucian education would think that it has added advantage of enabling one to identify, compare and rectify the shortcomings of one’s culture. They seemed to be making the best of the both worlds in their teaching. These findings shared much similarity with previous studies on language assessment done in Korea, Taiwan and Japan. For example, Kim and Kim (2017) reported a study on the varied effectiveness of instructor feedback (feeding-forward) in an integrated reading-to-write task for English based on the concept of the learning-oriented language assessment, an idea commonly known to have developed from the ‘West’. In Taiwan, Huang’s (2012) study on AfL L2 writing assessment found that teachers needed to make changes to their assessment practices by paying more attention to students’ performance and learning needs. The teachers also needed to know their students well that included their past learning history and beliefs, and their problems in learning to prioritise teaching. In addition, teachers also had to scaffold learning and made real-time decisions to provide useful feedback. It was found that these assessment changes were not overwhelming for these CHC teachers. In Japan, Saito and Inoi’s (2017) study on junior and senior high schools’ EL teachers showed that although there were individual differences in the use of FA, most Japanese teachers (80.49%) were found to be middle- to high-FA users. They went on to conclude that Japanese teachers were ready for FA. In other words, instead of treating new assessment ideas from the ‘West’ as barriers or are in conflict with the Confucian exam-oriented teaching, there are recent cases in the CHCs which reported harmonious alignment with the local education culture.

My research has extended beyond these past extant studies by examining teachers’ experience of AT in the context of a CHC that has, for many years, been labelled as a developed ‘East meets West’ nation-state. Although Singapore has such unique and idiosyncratic culture, it is also often considered to share certain cultural similarities with other CHCs. This research would provide important assessment lessons for other CHCs, as well as for the world.

The phenomenographical approach identifies the variation of the ways in which CL teachers experience and manage AT in this context, which were not investigated previously. This variation illustrates that the notion of CHCs may not be able to fully represent and explain the richness and nuances of all the CHCs educational jurisdictions including China, Hong Kong, Taiwan, Singapore, Japan, Korea, Macau and Vietnam. One

of the reasons is that there are various ways of conceptualising CHCs and in reality, the 'thickness' of such heritage culture also varied. For example, being the origin of the Confucius thinking, China has the 'thickest' essence of the CHCs. On the other hand, Hong Kong and Singapore, formally colonised by the British, are seen to be readier to adopt and apply constructivist principles of assessment from the 'West' than the mainland Chinese teachers.

As concluded by Leong and Tan (2014), Singapore's case of using FA has found many mediating influences of classroom assessment that include the importance of the sociocultural dimension. Accordingly, it may not be fruitful to expect the global theory of classroom assessment to be practised in the same way in different contexts. As they further explained:

Various elements of culture and politics can interplay in many and complex ways that are embedded in a national culture as a whole. Definitive generalisations are ideals at best, and any deep understanding should come only from case studies within individual countries. (Leong & Tan, 2014, p.615)

Singapore's case shows that the teachers' willingness to accept new assessment ideas from the 'West' is strong. This strength originated from the unique AT experiences that the teachers went through in the local context, which I have briefly described in Chapter 1. This research shows that the new concepts of assessment from the 'West' have really challenged some teachers' deeply-held conceptions of assessment, a similar view which was held by Black and Wiliam (1998) and Shepard (2001) in their studies. Indeed, the teachers in Conception C felt challenged when they must use assessment to prepare their students well for both the national examination, as well as lifelong learning. However, this AT allows teachers to make changes to what they traditionally believe classroom teaching and assessment should be.

All in all, CL teachers' experience and management of AT in this research allowed me to re-discover the open tradition of Confucian education and make predictions on how other CHCs may take up new assessment ideas from the 'West'. As Singapore moves into the next unpredictable educational era, teachers would need to be

adaptable and flexible to assessment changes, while still holding on to their exam-oriented practices. Such findings help to avoid the stereotyping of certain teachers' behaviour or thinking to a fixed frame of conclusion and explanations.

5.3 Implications and recommendations for practice and policy

In the same way as how Tan (2009) uses phenomenography to describe teachers' experience of educational policy, I argue that phenomenography is also useful for describing teachers' different ways of experiencing AT from the perspective of how it is experienced. Subsequently, this variation of experiencing tension in assessment may in turn be understood and have direct implications in the theoretical, analytical and pedagogical aspects. The table 5.3a below shows the three aspects with the inter-related questions and area of concern:

Aspects	Questions	Area of concern
Theoretical	What is a way of experiencing AT?	This concerns the theory of awareness, which is a key to understand how teachers may experience the phenomenon (AT) in distinctively different ways.
Analytical	What is the exact difference among the varied ways of experiencing AT?	This concerns the analytical structure of awareness in the outcome space. The analytical structure of awareness explains the differences in terms of the dimensions of variation among different conceptions, as well as the differences among categories of description.
Pedagogical	How can different ways of experiencing AT be achieved?	This concerns how the patterns of variation in experiencing AT in a specific context may be brought about for teachers to experience AT.

Table 5.3a: The three aspects of phenomenographic study

In short, these three aspects depict how the experience of AT is theorised by each individual teacher (theoretical aspect), analysed to separate the major difference as a way of experiencing AT and how this variation is related (analytical aspect) and finally to suggest how ‘better’ (or sophisticated) ways of experiencing AT may be brought about (pedagogical aspect).

The theoretical aspect posits that each teacher may experience AT in a distinctively different way. These conceptualisations and articulations by the teachers of what AT is and how it is managed from a small group of teachers may not voice out of the experiences of the majority of teachers who may not experience AT in very similar ways. An alternative to such a gap would be to theorise teachers’ distinctive AT experience from phenomenography. In this research, a structure of awareness of AT is derived from the nine CL teachers after the analysis was done.

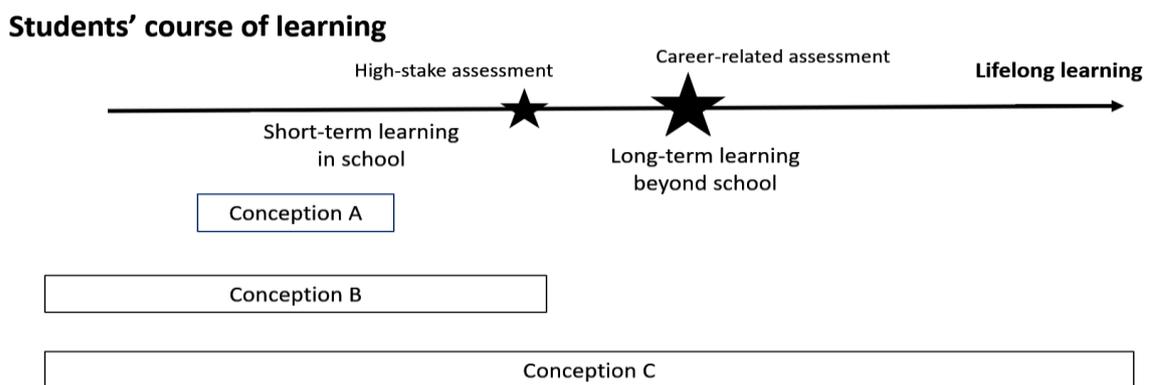


Figure 5.3: The structure of awareness of AT

The structure of awareness of AT conceptualises how CL teachers may experience AT in qualitatively different ways. It can explain the epistemology of experience of AT as it is understood and experienced by teachers. The structure of awareness in Figure 5.3 illustrates how each conception of AT are separated and yet linked. It allows readers to conceptualise how teachers may experience the phenomena of AT in different ways. Conception A’s focal awareness is on current limitation while Conception B is on imminent opportunities. Both conceptions pay more attention on students’ short-term learning in school. Conception C is most sophisticated for its focal awareness is on students’ long-term (or life-long) learning that is beyond school.

These conceptions form the “outcome space” and may be understood in terms of three dimensions (assessment purposes, reality and ideal gap, involvement of stakeholders of education). It is these dimensions of variations that explain how one category of description is different from the rest. This is the analytical aspect of the outcome space.

I argue that such approaches to investigating and understanding how AT is experienced is particularly useful in instances where AT has always been seen to be negative and undesirable to teaching (such as Conception A) in many educational systems. These ‘poorer’ (or less sophisticated) ways of experiencing AT could be possibly contributed by the top-down introduction of new teaching and assessment policies, which may directly bring about much tension to the school leaders. Consequently, school leaders demand that their teachers comply with these new policies and may require a change in their current assessment practices. At the end, teachers felt the tension introduced by these new policies. Such understanding and interpretation of how AT is experienced has not been helpful as it carries a high degree of negative connotation and there is no structure to understand and help teachers improve their experience of AT.

Alternatively, the varied ways of experiencing AT reminds policy-makers and school leaders to embrace these differences and reject stereotyping. Teachers’ experience in the less sophisticated conceptions implies that AT can be negative and undesirable. However, AT could also be constructive and useful for some teachers in the more sophisticated conceptions. This sophistication reminds policy-makers, school leaders and teachers not to avoid or pretend that AT is not at work because the consequence of ignoring this practice problem may bring about negative long-term effects to teaching and learning. One of the possible outcomes for ignoring AT is that some teachers would continue to struggle with assessment if policy-makers and school leaders fail to improve the situation to bring about ‘better’ ways of experiencing AT. In the same vein, if AT is ignored, the more sophisticated and constructive conceptions would hardly be brought about. School leaders would not be able to advance the ‘poorer’ conception to the ‘better’ ones. To this end, there should not be any compromise for school leaders and teachers when dealing with AT. This structure of awareness of AT will also be particularly useful for informing individuals to be more aware of the variation in

experiencing a phenomenon. The findings may allow others to understand, appreciate and respect teachers' different experience of AT. It alerts policy-makers and school leaders to avoid complacency and heighten their sensitivity towards teachers' distinctive experience of AT.

So far, I have theorised a way of experiencing AT, analysed it by breaking down its component parts for readers to learn about its structure of awareness of AT. More importantly, I shall suggest in the pedagogical aspect how 'better' ways of experiencing AT can be brought about and taught. The phenomenographic argument for using the structure of awareness to improve pedagogical experiences of AT would describe how some ways of experiencing tension in assessment are 'better' than others and suggest at different levels how more desired ways of experiencing AT can come about.

At the individual level, the findings allow teachers to learn to become more aware of variation in experiencing the phenomenon and use it to compare with their own ways of experiencing the same phenomenon. At the policy level, by understanding the collectively different ways of experiencing AT, policy-makers and school leaders may then have greater awareness and also share a common understanding of what it means to experience AT. To this end, future assessment practice, policy, reforms or initiatives and discussions regarding teaching training and professional development may be premised on these lived experiences reported. The pedagogical dimension of this phenomenographic research could inform readers on how the 'better' (or more sophisticated conceptions) ways of experiencing AT may be brought about and taught.

In phenomenography, the theory of learning considers learning to have happened when an individual's awareness of the experienced phenomenon has changed. This means that the individual must have experienced differences in ways of experiencing a phenomenon (Marton & Booth, 1997). To learn about how the more sophisticated conceptions can be brought about and taught, it is vital to consider the conditions or environment conducive to such 'better' experience. Earlier, Tsui (2004) pointed out three necessary tasks for this type of learning to happen. First, there must be a correct condition for learners to discern and be aware of the critical aspects of the learning object. Second, teachers should be aware of the learners' experience of the object. Third, teachers should spread the common understanding of how learners could experience the different awareness of the object of learning.

Similarly, the findings from this research could enhance our understanding of AT and help teachers better manage it. As explained, the variation of experiencing AT are defined by whether teachers' focal awareness is on students' short-term or long-term learning. Teachers experienced 'better' AT (such as Conception C) when their focal awareness and end point is on students' long-term learning. Likewise, teachers' experience of AT (such as Conception A) turned out to be 'poor' when their focal awareness and end point is on students' short-term learning and achievement. This structure of awareness of AT serves as a foundation for the pedagogical dimension of this phenomenographic research.

Focus	Short-term learning	Long-term learning
Teachers' focal awareness and end point	SA and national examination	National examination and beyond (lifelong learning)
Main learning goal	Do well in national examination	Prepare for lifelong learning
Awarded for	Achieving certain standards (such as language proficiency) within a limited period of learning	Acquiring certain learning attitude, disposition and habits over a long period of learning
Assessment focus	Subject knowledge and skills	Learning strategies and habits, cognitive and non-cognitive development, critical thinking and creativity
Characteristics of learning	Non-integrated, limited scope and periodical	Holistic, broad-based and on-going (lifelong)

Table 5.3b: Agendas between short- and long-term learning

Given the challenges and tension experienced by teachers in assessment works, how sophisticated should teachers' experience of AT be in order for more ambitious assessment work to be sustained in schools? I argue that one of the important enabling factors, which influence the ambition of assessment effort that will benefit quality learning lies in the contrasting agendas between working on students' short-term or long-term learning achievement. This argument is developed from the structure of awareness of AT discussed above. While in reality, short- and long-term learning may not be experienced in such a tidy and simple conceptualisation, Table 5.3b highlights the main agendas focus between short- and long-term learning. For example, while the main learning goal in the short-term learning context focuses on doing well in national examinations, the long-term learning would focus more on the preparation of students' lifelong capacity building, interest cultivation and their joy of learning. The assessment focus in the short-term learning context is on the subject knowledge and skills. In the case of language learning, this may include listening, speaking, reading, writing knowledge and skills. On the other hand, in the long-term learning context, the assessment would pay more attention to students' learning strategies and habits, cognitive and non-cognitive development, as well as critical thinking, creativity and interest. These aspects of students' learning would require a more holistic approach and much longer time to develop and assess.

To this end, teachers and school leaders may consider putting the long-term learning agendas above the short-term learning agendas so that teachers' experience of AT could be meaningfully envisaged to a more sophisticated conception. School leaders may also reconsider practices that encourage short-term decision-making in learning and assessment. This may include mid-year or annual placement test, banding of students according to examination results and academic award systems that are too frequently carried out. For the past decades, although the MOE has been promoting the use of MTL as a living language among the students, not all stakeholders of education buy into such a far and broad perspective on the learning of the MTL. Teachers, students and parents tend to pay more attention to the short-term CL learning focus, such as doing well in SA and the national examinations. To encourage teachers to focus on long-term and living language learning, future curriculum and assessment planning should emphasise and build in more opportunities to develop language skills that support

lifelong learning (such as finding students' interest in extensive reading and writing), rather than on the short-term language acquisition and its related assessment. Alternative assessment that focuses on students' efforts and good learning habits, as well as critical thinking and creativity could be encouraged in schools. To this end, policy-makers may consider a movement or initiative that requires curriculum planning and development of all subjects to focus their attention to cultivating students' long-term learning disposition, rather than short-term learning and achievement. Pre-service and in-service teacher education and professional training should also echo such focus. More importantly, policy-makers should also establish coherence from the policy to programmatic and eventually to school and classroom levels. School leaders and school middle managers should be well-informed of such movement or initiative and be clear about the instruction on how they can support such an intent into practice.

My recommendations have attracted significant interest from colleagues of other units in the Curriculum Planning and Development Division (CPDD) since I reported my research at a professional forum event. They found relevance in my research method, implications and recommendations for policy and practice and wanted to find out more from me.

To policy-makers, the research findings present the variation of teachers' experience of AT. On the one hand, the evidence from the conceptions raises issues that require policy-makers' attention, in that teachers indeed experienced different degree of AT in their teaching. Although some teachers experienced AT in 'better' ways, some still found it to be negative and detrimental to their teaching. These teachers experienced AT to be limiting their teachings and thus no changes were made to their assessment practice. This group of teachers would require more support from policy-makers in order to see fundamental changes in their assessment practices. More importantly, expectation or performativity that points to teachers' accountability in terms of attaining good schools and national examination results (the short-term learning perspective) needs reviewing. Otherwise, teachers will still experience AT from the new assessment initiative and the old belief that continues to encourage the "traditions" of working on delivering good national examination results. In addition, there should be coherence among all the different policies relevant to teachers' sound assessment practices that support students' long-term learning (or lifelong learning),

instead of emphasising short-term achievement. Policies at various educational levels and divisions, such as curriculum making and planning, teaching and assessment practice, schools and teachers' appraisal, educational technology, teaching training and professional development, higher education, as well as student development should be consistent in supporting students' long-term learning.

Policy-makers should also help all stakeholders of education to have a better understanding of the objectives, processes and products of different types of assessment in schools. They should be informed of the essential concepts and principles of different assessment (including educational measurement, feedback and so on). These would better inform them about the main difference between short- and long-term learning and their related assessment.

On the other hand, the more advanced conceptions are encouraging. Some Singapore teachers are able to manage assessment challenges and tension well. For these teachers, policy-makers should push for policy initiatives which would allow more room for ambitious assessment endeavour. These may include the use of learning-oriented assessment (Carless, 2006, 2009; Jones, Saville & Salamoura, 2016), dynamic assessment (Poehner, 2008; Poehner & Infante, 2017) and sustainable assessment (Boud, 2000; Boud & Soler, 2016). Although some of the findings are optimistic in this research, assessment policy changes would remain challenging in future. One of the possible reasons is that teachers' comprehension and interpretation of the policy remain diverse and messy due to a lack of clarity and directional instruction on how they could realize the policy intent into practice. While there may be a need to inform teachers about the philosophy and theory behind any new assessment initiative or change, it is more important for policy-makers to be more explicit about the implementation of the initiative in classroom teaching and learning. Particularly, it would be beneficial for policy-makers to define the different roles in the implementational process. Last but not least, the variation of teachers' experience of AT also reminds policy-makers that there is "an intermediate point" where teachers would have to go through before the ideal stage (such as Conception C) is reached in any assessment reform or initiative. Policy-makers, school leaders and teachers will need to invest time, effort and stay patient to see the changes in practice.

5.4 Limitations of the study and suggestions for further research

Despite careful considerations being paid to the methodology of the study in terms of validity, reliability and “bracketing” issues, there are several limitations which are acknowledged in this section.

Due to time, resource and contextual constraints, this research focused only on the CL teachers, whom to a large extent, could represent a CHC in Southeast Asian countries. Focusing solely on CL teachers implies a limited disciplinary scope. In the extant research, some studies have focused only on literacy and numeracy teachers (such as Kostogriz & Doecke, 2011; Dreher, 2012; Hardy, 2013; Suurtamm & Koch, 2014), while others investigated on Sciences or the STEM teachers (Kim et al., 2013; Paechter, 1995; Tan and Leong, 2014). There is also another group (Bishop et al., 1999; Lindsay & Lewis, 2003; Tan, 2013; Volante & Beckett, 2011) who analysed teachers’ assessment experiences in multiple subjects. Other studies have explored teachers’ assessment experiences in higher education (Bonner & Chen, 2009; Collins et al., 2010; Ecclestone & Swan, 1999; Harman & McDowell, 2011; Hornby, 2003; Reis & Villaume, 2002; Scholtz, 2007). Many also have examined the experience and response of teachers when new assessment reforms or initiatives are introduced (Black, 1993; Crooks, 1998; Blok, Otter & Roeleveld, 2002; Cohen & Ball, 1990; Hogan, 2011; Vandeyar, 2005). All these studies could be further improved with more participants. This would allow a comparison across different subjects. Subsequently, teachers’ experiences of AT gleaned from this research may be used to compare if they are peculiar to CL teachers or are applicable to other subjects.

In this phenomenographic research, I was not able to guarantee that all utterances receive an equal amount of attention. During data analysis, while the themes and categories of description might have been constructed through an iterative process, there were variations to the use of the data. While some were used to build the conceptions, others were used for verification of the conceptions. Future research may use other research approaches that could provide new perspectives on teachers’ experience of AT.

Another limitation of this study is the extent in which the findings may be transferred to other contexts. According to Marton and Booth (1997), the extent to

which the findings can be applied in other contexts in phenomenographic study is mostly left to the consumer. There may be a need to replicate future study on this topic in other contexts since Singapore has a unique and idiosyncratic culture.

Although care has been taken of in terms of purposeful selection of participants so that the spread of characteristics could be used to maximize the “conceptual variations” in the data, some CL teachers may not be represented in the selection. These include teachers with special backgrounds from Taiwan or Mainland China with teaching experience of more than 20 years.

Lastly, as explained in Chapter 3, to truly capture the authentic voices from the teachers and to avoid linguistic barrier, Mandarin Chinese has been used to communicate during the interviews. Care has been taken to ensure the accuracy during translation. However, some of the language expressed may still be quite different between the East and the West, mainly due to cultural difference. For example, the translated utterances may seem strange for a native English reader but could be quite normal for the CL teachers.

Possibilities for future research should focus on how teachers in other specific cultural contexts, experience AT when a new assessment idea is introduced. Because of the number of phenomenographic studies that focus on teachers’ experience of AT in a unique CHC, it should be possible to juxtapose these experiences, with those that are in other cultural contexts. Such future studies would confirm similarities, but more importantly point out critical differences and thus make comparisons (and perhaps predictions) on how teachers in various cultural contexts experience and manage AT. Another possibility is to research on different types of language teacher in terms of acquisition (first, second, third, foreign or heritage language). This would allow comparisons among different language teachers in the ways they experience and manage AT.

5.5 My experience of phenomenographic research

To start with, I had no previous experience with phenomenographic research, though I was familiar with the basic educational research methods. My first research proposal was a mixed-method. I started off with my interest in teachers’ conceptions of assessment research. I found that most of these past studies used quantitative approach,

instead of qualitative. Then I also observed that many assessment studies mentioned the tension teachers experienced when carrying out assessment. Subsequently, I became interested to investigate their different experiences and their ways of dealing with AT, which I think is a practice problem that current scholarship has not been able to address satisfactorily. In order to solve this practice problem, I first needed to know the description of the collective variations of teachers' conceptions of the experienced phenomenon.

Based on my research interest and the literature review, phenomenography was chosen due to its ability to reveal teachers' varied experiences and thus the conceptions of AT. My journey with phenomenography mirrored that of Tan (2012, 2013), in his study of Singapore primary school teachers' experiences of alternative assessment.

As I began to reflect on my experience with this methodology, several observations were made. The first challenge was on the crafting of the interview questions as different types of question served different purposes in varied stages. As reported in Chapter 3, I selected five types of questions for the interviews. Although pilots were conducted, I still struggled to employ the right types of question at the beginning stage. As Dörnyei (2007, p.140) pointed out the features of a good interview: (a) it flows naturally, and (b) it is rich in detail, my earlier interviews were neither natural nor rich. However, after a few rounds of practice and experimentation, I became more natural in asking the right questions at the appropriate time. My conversations with the teachers also became richer and more in-depth as I progressed.

The second challenge pertains to my experience as an educator and the issue of ensuring the validity and reliability of this research. To this end, "bracketing" becomes paramount importance because I needed to be accountable and able to communicate to the readers of the steps and efforts taken in the phenomenographic analysis that included arriving at the categories of description and the outcome space. Ashworth (1999) explains that the purpose of "bracketing" is: "to set aside theories, research presuppositions, ready-made interpretations etc., in order to reveal engaged, lived experience" (p. 707). This proven to be tough because I was supposed to describe the teachers' own experience of AT rather than my own experience interacting with them. This was especially challenging because of my rich experiences as an educator that included being a classroom language teacher, middle manager, curriculum planner,

subject specialist, as well as an assessment consultant. This broad knowledge of the language teaching, as well as my frequent interactions with CL teachers made “bracketing” very difficult. However, after many iterative reviews of the data and categories of description, along with the inputs and checks from other phenomenographic experts, I began to confidently confirm my categories of description (conceptions of AT) and eventually the outcome space.

Overall, the most challenging issue was to ensure the rigor and quality of this research as there were no common views within the qualitative research community on the criteria for checking the quality of phenomenographic studies. For example, how do I justify the utterances from the interview, the conceptions and the outcome space are validly linked? How do I interpret the linguistic differences and choice of words among the nine teachers and ensure that they mean the same thing, especially when I need to translate the interview data from Mandarin Chinese to English? I had to take on these challenges one by one. To ensure that the interview data, the codes and the conceptions were linked in a valid way, I had encouraged participants to reflect on the intended meaning of what they said when it was necessary, a method supported by phenomenographic researchers (such as Anderberg, 2000; Sin, 2010). As for the linguistic and translation issues, experts in these two specialised areas were consulted, in addition to member’s checking. These tasks were time-consuming but essential to ensure research quality. Despite all these efforts, the subjective nature of phenomenographic studies could not be totally eliminated. The many ways of reading and interpretation of the categories of description and outcome space by the readers were normal.

In closing, from my experience and reflection of using phenomenography as a research method, I concluded that it was a valid and useful approach to investigate an experienced phenomenon, as long as one had spent enough effort working through issues such as those described above. More importantly, this research had also heightened the AT I experienced as a teacher. Without this experience, I would still believe that AT is always negative and undesirable for teaching and learning.

5.6 Summary

This final chapter concludes my entire research on CL teachers' experiences of AT. I have recapitulated the research purpose and findings, reflected on some of the limitations, before going into detail on the significance and contributions of the research in the four main areas.

With the findings, I discussed the implications and made some recommendations for policy, research and practice in schools. Before ending this chapter, I reflected on my experience of doing phenomenographic research.

I hope that the findings and ideas presented in this research will provide an alternative way for educators to better understand how teachers are experiencing AT in reality, a phenomenon that has been reported to be very common in all education jurisdictions but was poorly understood. I also hope that the implications and recommendations for future assessment practice and policy, as well as research will prove particularly useful.

References

- Akerlind G.S. (2005). Variation and Commonalities in Phenomenographic Research Methods. *Higher Education Research & Development*, 24(4), 321-334.
- Akerlind, G.S. (2008). A phenomenographic approach to developing academics' understanding of the nature of teaching and learning. *Teaching in Higher Education*, 13(6), 633-644.
- Alkharusi, H. (2011). Development and Datametric Properties of a Scale Measuring Students' Perceptions of the Classroom Assessment Environment. *International Journal of Instruction*, 4(1), 105-120.
- Alkharusi, H. (2013). Canonical Correlational Models of Students' Perceptions of Assessment Tasks, Motivational Orientations, and Learning Strategies. *Journal of Instruction*, 6(1), 21-38.
- Anderberg, E. (2000). Word meaning and conceptions: An empirical study of relationships between students' thinking and use of language when reasoning about a problem. *Instructional Science*, 28 (2), 89–113.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3), <http://PAREonline.net/getvn.asp?v=10&n=3>
- Andrade, H., Du, Y., & Wang, X. (2008). Putting Rubrics to the Test: The Effect of a Model, Criteria Generation, and Rubric-Referenced Self-Assessment on Elementary School Students' Writing. *Educational Measurement: Issues And Practice*, 27(2), 3-13. <http://dx.doi.org/10.1111/j.1745-3992.2008.00118.x>
- Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing*, 21(3), 284–304.
- Ashworth, P. (1999). "Bracketing" in phenomenology: Renouncing assumptions in hearing about student cheating. *International Journal of Qualitative Studies in Education*, 12 (6), 707-721.
- Ashworth, P., & Lucas, U. (2000). Achieving empathy and engagement: A practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education*, 25(3), 295–308.
- Avidov-Ungar, O., Friedman, I. & Olshtain, E. (2014). Empowerment amongst teachers holding leadership positions. *Teachers and Teaching: theory and practice*, 20(6), 704-720.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University.

- Bachman, L. F. (2000). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Barab, S. A., Evans, M. A., & Baek, Eun-Ok. (2004). Activity Theory as a Lens for Characterising the Participatory Unit. In D. H. Johassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp.199-214). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Barrand, J. L., & Rettig, R. (2011). An Investigation of Teacher Training and Perceptions Regarding Reading Assessment with Elementary Students. *ERIC Online journal*, <http://files.eric.ed.gov/fulltext/ED523329.pdf>
- Bassey, M. (1999). *Case study research in educational settings*. Maidenhead, UK: Open University Press.
- Baumfield, V. (2016). Practitioner Research: Understanding Teaching Through Inquiry. In D. Wyse, L. Hayward & J. Pandya, *The SAGE Handbook of Curriculum, Pedagogy and Assessment* (pp. 154-168). London: SAGE Publication Ltd.
- Bennett, M. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10 (2), 179-195.
- Bennett, M. (2004). Becoming interculturally competent. In J. Wurzel (Ed.), *Toward multiculturalism: A reader in multicultural education* (2nd ed., pp. 62-77). Newton, MA: Intercultural Resource
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5-25.
- Benson, P. (2010). Teacher education and teacher autonomy: Creating spaces for experimentation in secondary school English language teaching. *Language Teaching Research*, 14(3), 259-275. <http://dx.doi.org/10.1177/1362168810365236>
- Berry, A. (2007). Reconceptualizing Teacher Educator Knowledge as Tensions: Exploring the tension between valuing and reconstructing experience. *Studying Teacher Education*, 3(2), 117-134. doi: 10.1080/17425960701656510
- Berry, R., & Adamson, B. (2011). *Assessment reform in education: policy and practice*. Dordrecht: Springer.
- Biggs, J. (1996). Western misperceptions of the Confucian-heritage learning culture. In D. Watkins & J. Biggs (Eds.), *The Chinese Learner: Cultural, Psychological and Contextual Influences* (p.45–68). Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.
- Biggs, J. (1998). Assessment and Classroom Learning: a role for summative assessment?. *Assessment In Education: Principles, Policy & Practice*, 5(1), 103-110.

- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., & Looney, A. et al. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures In Education*, 13(1), 117-140. <http://dx.doi.org/10.1177/1478210314566733>
- Bishop, K., Bullock, K., Martin, S., & Thompson, J. (1999). Users' perceptions of the GCSE. *Educational Research*, 41(1), 35-49.
- Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment In Education: Principles, Policy & Practice*, 22(1), 161-177. <http://dx.doi.org/10.1080/0969594x.2014.999643>.
- Black, P., & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: GL Assessment.
- Black, P., & Wiliam, D. (1998c). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P. (1995). Curriculum and assessment in science education: the policy interface. *International Journal of Science Education*, 17(4), 453-469.
- Black, P. (1998). *Testing: Friend or Foe?* London: Falmer.
- Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21(1), 49-97.
- Black, P., & Wiliam, D. (2003). 'In praise of educational research': formative assessment. *British Educational Research Journal*, 29(5), 623-637. <http://dx.doi.org/10.1080/0141192032000133721>
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). London: Sage.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice*. Berkshire: Open University Press.
- Blaikie, N. (2000). *Designing Social Research*, Cambridge: Polity.
- Blok, H., Otter, M. E., & Roeleveld, J. (2002). Coping with conflicting demands: Student assessment in Dutch primary schools. *Studies in Educational Evaluation*, 28(2), 177-188.
- Bloom, B. S. (1970). Toward a theory of testing which includes measurement, evaluation, and assessment. In M. C. Wittrock & D. E. Wiley (Eds.), *The evaluation of instruction*. New York, NY: Holt, Rinehart, & Winston.
- Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods*. New York: Pearson Education group.

- Bogler, R., & Somech, A. (2004). Influence of teacher empowerment on teacher organizational commitment, professional commitment and organizational citizenship behavior in schools. *Teaching and Teacher Education*, 20(3), 277-289.
- Bonner, S. M. (2016). Teachers' perceptions about assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 21–39). New York: Routledge.
- Bonner, S. M., & Chen, P. P. (2009). Teacher Candidates' Perceptions About Grading and Constructivist Teaching. *Educational Assessment*, 14(2), 57–77.
- Booth, S. (1992). *Learning to program: A phenomenographic perspective*. Goteburg: Acta Universitatis Gothoburgensis.
- Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long - term learning. *Assessment & Evaluation In Higher Education*, 31(4), 399-413. <http://dx.doi.org/10.1080/02602930600679050>
- Bowden, J. (2000). The nature of phenomengraphic research. In J. Bowden & E. Walsh (Eds.), *Phenomenography* (pp.1-18). Melbourne: RMIT Publishing.
- Bowden, J. (2005). Reflections on the phenomenographic team research process. In J. Bowden & P. Green (Eds.), *Doing developmental phenomenography* (1st ed.) (pp.11-31). Melbourne: RMIT University Press.
- Bowe, R., Ball, S. J., & Gold, A. (1992). *Reforming education and changing schools: Case studies in policy sociology*. London: Routledge.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge: Cambridge University Press.
- Broadfoot, P. (1996). *Education, Assessment and Society: A Sociological Analysis*. Buckingham: Open University Press.
- Broadfoot, P. (2002). Beware the Consequences of Assessment!. *Assessment In Education: Principles, Policy & Practice*, 9(3), 285-288. <http://dx.doi.org/10.1080/0969594022000027645>
- Broadfoot, P. (2007). *An introduction to assessment*. New York: Continuum.
- Brown, G. T. L., & Harris, L. R. (2012). Student conceptions of assessment by level of schooling: Further evidence for ecological rationality in belief systems. *Australian Journal of Educational & Developmental Psychology*, 12, 46-59.
- Brown, G. T. L. (2003). Teachers' instructional conceptions: Assessments' relationship to learning, teaching, curriculum and teacher efficacy. Paper presented at the joint

conference of the Australian and New Zealand Associations for Research in Education (AARE/NZARE), Auckland, 28 Nov.-3 Dec.

- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validations of an abridged version. *Psychological Reports*, 99, 161-170.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45-70.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and mathematics achievement: Evidence for the power of self-regulation. *Australian Journal of Educational and Developmental Psychology*, 7, 63-74.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75-89.
- Brown, G. T. L., Hui, S. K. F., Yu, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50(5-6), 307-320.
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice*, 16(3), 347-363.
- Brown, G. T. L., Lake, R., & Matters, G. (2011). New Zealand and Queensland teachers' conceptions of curriculum: Potential jurisdictional effects of curriculum policy and implementation. *Curriculum Perspectives*, 31(3), 33-48.
- Brown, J. D. (2004). Performance assessment: existing literature and directions for research. *Second Language Studies*, 22(2), 91-139.
- Brown, M. (2011). Going back or going forward? Tensions in the formulation of a new National Curriculum in mathematics. *The Curriculum Journal*, 22(2), 151-165.
- Bucci, P. (2002). *Teacher knowledge, beliefs and practices of classroom assessment: from the perspective of five elementary teachers*. Thesis (Ed.D)--University of Toronto, 2002.
- Bullock, D. (2010). Learner self-assessment: an investigation into teachers' beliefs. *ELT Journal*, 65(2), 114-125.
- Byrne, J., Downey, C., & Souza, A. (2013). Teaching and learning in a competence-based curriculum: the case of four secondary schools in England. *The Curriculum Journal*, 24(3), 351-368.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Canton, J. (2006). *The extreme future : the top trends that will reshape the world for the next 20 years*. USA: Plume (Penguin Group).
- Carless, D. (2006). *How Assessment Supports Learning: Learning-oriented Assessment in Action*. Hong Kong University Press, HKU.
- Carless, D. (2009). Learning-oriented assessment: Principles, practice and a project. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P.M. Johnston, & M. Rees (Eds.), *Tertiary Assessment & Higher Education Student Outcomes: Policy, Practice & Research* (pp.79-90). Wellington, New Zealand: Ako Aotearoa.
- Carless, D. (2011). *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. New York: Routledge.
- Carless, D., & Lam, R. (2014). Developing assessment for productive learning in confucian-influenced settings. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning (Vol. 1)* (pp. 167–179). Dordrecht, Netherlands: Springer
- Chai, C. S., Chin, C. K., Koh, J. H. L., & Tan, C. L. (2013). Exploring Singaporean Chinese Language Teachers' Technological Pedagogical Content Knowledge and its Relationship to the Teachers' Pedagogical Beliefs. *The Asia-Pacific Education Researchers*, 22(4), 657-666.
- Chamberlin, M. T., Farmer .J. D., & Novak, J. D. (2008). Teachers' perceptions of assessments of their mathematical knowledge in a professional development course. *Journal of Mathematics Teacher Education*, 11(6), 435-457.
- Chappuis, S., Commodore, C., & Stiggins, R. (2017). *Balanced assessment systems*. California: Corwin.
- Chan, K. T. (2010). *Song Dai Zhuang Yuan Yan Jiu (A Study of Top Scholars (Zhuang Yuan) in the Chinese Civil Examinations of Song Dynasty)* (Unpublished master's thesis). National University of Singapore, Singapore.
- Chan, K. W., & Elliott, R. G. (2004). Relational analysis of personal epistemology and conceptions about teaching and learning. *Teaching and Teacher Education*, 20, 817-831.
- Chang, Chiung-Wen. (2006). Teachers' Beliefs towards Oral Language Assessment in Taiwan Collegiate EFL Classrooms. Paper presented at 2006 International Conference on English Instruction and Assessment. National Chung Cheng University, Chia-Yi, Taiwan.
- Chappuis, J. (2009). *Seven strategies of assessment for learning*. Princeton, N.J.: Educational Testing Service.

- Chappuis, J. (2012). *Classroom assessment for student learning: doing it right, using it well*. Upper Saddle River, N.J. : Pearson.
- Chappuis, S., Commodore, C., & Stiggins, R. (2010). *Assessment balance and quality: an action guide for school leaders (3rd ed.)*. Boston: Pearson Education, Inc.
- Chappuis, S., Commodore, C., & Stiggins, R. (2017). *Balanced assessment systems*. California: Corwin.
- Cheah, Y. M. (1998). The Examination Culture and its Impact on Literacy Innovations: the Case of Singapore. *Language and Education*, 12(3), 192-209.
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, 21, 271–285.
- Cheng, K. M. (1990). The culture of schooling in East Asia. In N. Entwistle (Ed.), *Handbook of educational ideas and practices* (pp.163-173). London: Routledge.
- Cheng, Liying, Andrews, S., & Yu, Ying. (2010). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221-249.
- Choi, J., Kushner, K., Mill, J., & Lai, D. (2012). Understanding the Language, the Culture, and the Experience: Translation in Cross-Cultural Research. *International Journal Of Qualitative Methods*, 11(5), 652-665. doi: 10.1177/160940691201100508
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311-329.
- Cohen, D.K., & Ball, D.L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis*, 12(3), 233–239.
- Cohen, L., & Manion, L. (1989) *Research Methods in Education*, 3rd edn. London: Routledge.
- Collier-Reed, B. I., Ingerman, A., & Berglund, A. (2009). Reflections on trustworthiness in phenomenographic research: Recognising purpose, context and change in the process of research. *Education as Change*, 13(2), 339-355.
- Corney, G. (2006). Education for Sustainable Development: An Empirical Study of the Tensions and Challenges Faced by Geography Student Teachers. *International Research in Geographical and Environmental Education*, 15(3), 224-240.
- Creswell, J. W. (2012). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education.
- Creswell, J.W., & Plano Clark, V.L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks: Sage Publications.

- Creswell, J.W. (2007). *Qualitative inquiry & research design: choosing among five approaches*. Thousand Oaks: Sage Publications.
- Crook, C., Gross, H., & Dymott, R. (2006). Assessment relationships in higher education: the tension of process and practice. *British Educational Research Journal*, 32(1), 95-114.
- Crooks, T. J. (1998). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Cruickshank, K., & Tsang, L. (2011). Teaching and Learning Chinese: A research Agenda. In L.Tsung and K.Cruickshank (Eds.), *Learning and Teaching Chinese in Global Contexts: Multimodality and literacy in the new media age*. London: Continuum.
- Dahlin, B. (1999). Ways of coming to understand: Metacognitive awareness among first-year university students. *Scandinavian Journal of Educational Research*, 43, 191-208.
- Dahlin, B., & Watkins, D. (2000). The role of repetition in the processes of memorising and understanding: A comparison of the views of German and Chinese secondary school students in Hong Kong. *British Journal of Educational Psychology*, 70, 65-84.
- Dall'alba, G. (2000). Reflections on some faces of phenomenography. In J. Bowden & E. Walsh (Eds.), *Phenomenography* (pp.83-101). Melbourne: RMIT Publishing.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- Davison, C., & Hamp-Lyons, L. (2010). The Hong Kong Certificate of Education: School-based assessment reform in Hong Kong. In L-y Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 248–266). New York, NY: Routledge
- de Vaus, D. A. (2001). *Research design in social research*. Thousand Oaks, CA: SAGE.
- Delandshere, G. (1996). From Static and Prescribed to Dynamic and Principled Assessment of Teaching. *The Elementary School Journal*, 97(2), 105-120.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- Deneen, C. C., & Brown, G. T. L. (2011). The persistence of vision: An analysis of continuity and change in conceptions of assessment within a teacher education program. Paper presented at the 2011 International Association for Educational Assessment (IAEA) conference in Manila, Philippines.
- Deng, Z. (2010). Curriculum Planning and Systems Change. In Penelope Peterson, Eva Baker, Barry McGaw, (Eds), *International Encyclopedia of Education. volume 1* (pp. 384-389). Oxford: Elsevier.

- Deng, C., & Carless, D. R. (2010). Examination Preparation or Effective Teaching: Conflicting Priorities in the Implementation of a Pedagogic Innovation. *Language Assessment Quarterly*, 7(4), 285-302.
- Denzin, N. K., & Lincoln, Y. S. (2008). *Collecting and interpreting qualitative materials* (3rd Ed.). Thousand Oaks, CA: Sage Publication.
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203–214.
- Dewey, J. (1904). The relation of theory to practice in education. In *The Third yearbook of the National Society for the Scientific Study of Education: Part I: The Relation of theory to practice in the education of teachers* (pp. 9-30). Chicago: University of Chicago Press.
- Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development. An International Journal of Teachers' Professional Development*, 13(2), 173-186.
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.
- Doyle, W. (1992). Curriculum and pedagogy. In Jackson, P. W. (ed.) *Handbook of Research on Curriculum* (pp. 486–516). New York: Macmillan.
- Dreher, K. (2012). Tests, testing time and literacy teaching. *Australia Journal of Language And Literacy*, 35(3), 334-352.
- Ecclestone, K., & Swann, J. (1999). Litigation and Learning: Tensions in improving university lecturers' assessment practice. *Assessment In Education: Principles, Policy & Practice*, 6(3), 377-389. <http://dx.doi.org/10.1080/09695949992801>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University.
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.
- Entwistle, N. (1997). Introduction: Phenomenography in higher education. *Higher Education Research and Development*, 16(2), 127-134.
- Eren, A. (2010). Consonance and dissonance between Turkish prospective teachers' values and practices: Conceptions about teaching, learning and assessment. *Australian Journal of Teacher Education*, 35 (3), 27–48.
- Escudier, M. P., Newton, T. J., Cox, M. J., Reynolds, P. A. & Odell, E. W. (2011). University students' attainment and perceptions of computer delivered assessment: a comparison between computer-based and traditional tests in a 'high-stakes' examination. *Journal of Computer Assisted Learning*, 27(5), 440–447.

- Evans, L. (2002). 'What is Teacher Development?'. *Oxford Review of Education*, 28(1), 123-137.
- Ewell, P. T. (1987). *Assessment, accountability and improvement: Managing the contradiction*. Washington, DC: American Association for Higher Education.
- Ewell, P. T. (2008). Assessment and accountability in America today: Background and context. In V. M. H. Borden & G. Pike (Eds), *Assessing and accounting for student learning: Beyond the Spellings Commission (New Directions for Institutional Research, Assessment Supplement 2007)* (pp. 7–18). San Francisco, CA: Jossey-Bass.
- Ewell, P. T. (2009). *Assessment, accountability and improvement: Revisiting the tension*. Occasional Paper #1. National Institute for Learning Outcomes Assessment.
- Exley, B., & Chan, M. Y. (2014). Tensions between policy and practice : reconciliation agendas in the Australian curriculum English. *English Teaching: Practice and Critique*, 13(1), 55-75.
- Farrell, T. C., & Tan, S. K. K. (2008). Language Policy, Language Teachers' Beliefs, and Classroom Practices. *Applied Linguistics*, 29(3), 381-403.
- Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36(1), 51-62.
- Field, J. C. (1991). Educators' Perspectives on Assessment: Tensions, Contradictions and Dilemmas. *Canadian Journal of Education*, 16(2), 210-214.
- Finlay, L. (2002) Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research*, 2(2), 209-230.
- Firestone, W. A. (1998). A tale of two tests: Tensions in assessment policy. *Assessment in Education*, 5(2), 175-191.
- Fleer, M. (2015). Developing an assessment pedagogy: the tensions and struggles in re-theorising assessment from a cultural–historical perspective. *Assessment In Education: Principles, Policy & Practice*, 22(2), 224-246. <http://dx.doi.org/10.1080/0969594x.2015.1015403>
- Fleming, D., Bangou, F. & Fellus, O. (2011). ESL teacher-candidates' beliefs about language. *TESL Canada Journal*, 29(1), 39-56.
- Fletcher, R.B., Meyer, L.H., Anderson, H., Johnston, P., & Rees, M.T. (2012). Faculty and student conceptions of assessment in higher education. *Higher Education*, 64, 119–133.
- Flores, M. A. (2005). Teachers' views on recent curriculum changes: tensions and challenges. *The Curriculum Journal*, 16(3), 401-413.
- Fu, G. (2010). Speech by Ms Grace Fu, Senior Minister of State, Ministry of National Development and Ministry of Education, at the PERI Holistic Assessment Seminar 2010 on Tuesday, 13 July 2010 at 9.00 am at Agora Hall, Republic Polytechnic.

Ministry of Education Press Release. Retrieved from <http://www.moe.gov.sg/media/speeches/2010/07/13/peri-holistic-assessment-seminar-2010.php>

- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco, Calif.: Jossey-Bass.
- Gao, Ling-biao, & Kennedy, J. K. (2011). Zhong guo nei di jiao shi de kao ping guan: lei xing te dian yu qi shi (Teachers' conception of assessment in Mainland China: a3 categories, characteristics and implications). *Jiao yu xue bao (Journal of Educational Studies)*. 2011(1), 39-47.
- Gao Ling-biao, & Watkins, D. (2001). Identifying and assessing the conceptions of teaching of secondary school physics teachers in China. *British Journal of Educational Psychology*, 71, 443-469.
- Gareis, C. (2007). Reclaiming an Important Teacher Competency: The Lost Art of Formative Assessment. *Journal Of Personnel Evaluation In Education*, 20(1-2), 17-20.
- Geisinger, K. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304-312. <http://dx.doi.org/10.1037//1040-3590.6.4.304>
- George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge: MIT Press.
- Genc, Z. (2010). Teacher autonomy through reflective journals among teachers of English as a foreign language in Turkey. *Teacher Development*, 14(3), 397-409. <http://dx.doi.org/10.1080/13664530.2010.504028>
- Gilles, J., Detroz, P., & Blais, J. (2011) An international online survey of the practices and perceptions of higher education professors with respect to the assessment of learning in the classroom. *Assessment & Evaluation in Higher Education*, 36(6), 719–733.
- Gioka, O. (2008). Teacher or Examiner? The Tensions between Formative and Summative Assessment in the Case of Science Coursework. *Research In Science Education*, 39(4), 411-428. doi: 10.1007/s11165-008-9086-9
- Gipps, C. (1994). *Beyond testing: towards a theory of educational assessment*. London: Falmer Press.
- Goh, C. T. (1997) *Shaping our Future: Thinking Schools, Learning Nation* (Ministry of Education, Singapore). <http://www.moe.gov.sg/media/speeches/1997/020697.htm>

- Goh, C. T. (1997). Speech by Goh Chok Tong on Thinking Schools, Learning Nation. Singapore: Ministry of Education.
- Goh, K. C. (2009). *A longitudinal study of teacher preparation and professional development: pre-service to beginning teacher*. Singapore: Office of Academic Quality Management, National Institute of Education.
- Goh, Y. S (2010). Teaching Chinese as a Second Language in the Multi-lingual society of Singapore. In International Society For Chinese Language Teaching (Eds.). *A selection of Papers presented at the 9th International Conference on Teaching Chinese as an International Conference* (pp. 80-87). Beijing: Gaodeng Jiaoyu Chubanshe [Higher Education Press].
- Goh, Y. S. (2010). *The Globalization of Chinese: A Singapore Perspective*. Beijing: The Commercial Press.
- Goodwin, A. L. (2010). Globalization and the preparation of quality teachers: rethinking knowledge domains for teaching. *Teaching Education, 21*(1), 19-32.
- Gopinathan, S. (1997). Educational development in Singapore: Connecting the national, regional and the global. *The Australian Educational Researcher, 24*(1), 1-12. <http://dx.doi.org/10.1007/bf03219637>
- Gopinathan, S. (2007). Globalisation, the Singapore developmental state and education policy: a thesis revisited. *Globalisation, Societies and Education, 5*(1), 53-70.
- Gopinathan, S. (2009) Educating the Next Generation. In B.Welsh, J.Chin & T. H.Tan. (Eds.), *Impression of the Goh Chok Tong years in Singapore* (pp. 240-251). Singapore: NUS Press & Co-published with the Institute of Policy Studies, Lee Kuan Yew School of Public Policy.
- Gordon, S., & Rees, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership, 7*, 345-368.
- Goulah, J. (2010). Francis W. Parker, Tsunesaburo Makiguchi, and Lucy Sprague Mitchell: Young geographers, geography studies, and community Life. *Schools: Studies in Education, 7*(1), 41-46.
- Grix, J. (2002). Introducing Students to the Generic Terminology of Social Research. *Politics, 22*(3), 175-186.
- Gronlund, N. (2006). *Assessment of student achievement (8th ed.)*. Boston: Pearson.
- Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school? *Assessment in Education: Principles Policy & Practice, 21*, 286-305.
- Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Eds.), *The paradigm dialog*, (pp.17-30). Newbury Park, CA: Sage.

- Gurr, D., & Drysdale, L. (2012). Tensions and dilemmas in leading Australia's schools. *School Leadership & Management: Formerly School Organisation*, 32(5), 403-420.
- Guskey, T. R., & Huberman, M. (1995). *Professional Development in Education: new paradigms and practices*. New York, Teachers College Press.
- Guskey, T. R., & Sparks, D. (1996). Exploring the relationship between staff development and improvements in student learning. *Journal of Staff Development*, 17(4), pp. 34-38.
- Guskey, T. R. (2000) *Evaluating Professional Development*. Thousand Oaks, CA, Corwin Press.
- Hairon, S., Chua, S.M.J., & Tan, C. (2007). *Understanding quantitative and qualitative research methodologies. Knowledge & inquiry: an introduction to research skills*. Singapore: Pearson Hall.
- Hamilton, L., Stecher, B., & Yuan, K. (2012). Standards-Based Accountability in the United States. *Education Inquiry*, 3(2), 149-170. <http://dx.doi.org/10.3402/edui.v3i2.22025>
- Hamp-Lyons, L., & Davison, C. (2010). *Longitudinal study on the schools-based assessment component of the 2007 Hong Kong Certificate of Education (HKCE) English Language Examination*. Hong Kong: Hong Kong Examinations and Assessment Authority/University of Hong Kong.
- Han, M., & Yang, X. (2001). Educational Assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*, 8(1), 5-10.
- Hardy, I. (2013). Testing that counts: Contesting national literacy assessment policy in complex schooling settings. *Australia Journal of Language And Literacy*, 35(3), 67-77.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and Learning*. London: SAGE Publications Ltd.
- Harlen, W., & Deakin Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning. Research evidence in education library*. London: EPPI-Centre Social Research Unit, Institute of Education.
- Harlen, W., & James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment In Education: Principles, Policy & Practice*, 4(3), 365-379. <http://dx.doi.org/10.1080/0969594970040304>
- Harris, L., & Brown, G. (2016). Volume Introduction: The human and social experience of assessment: valuing the person and context. In G, Brown & L, Harris (Eds.), *Handbook of human and social conditions in assessment (1st ed.)*. New York: Routledge.

- Hasselgren, B. & Beach, D. (1997). Phenomenography — a “good for nothing brother” of phenomenology? Outline of an analysis. *Higher Education Research & Development*, 16(2), 191-202.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany: State University of New York Press.
- He, M. (2013). East~West Epistemological Convergence of Humanism in Language, Identity, and Education: Confucius~Makiguchi~Dewey. *Journal Of Language, Identity & Education*, 12(1), 61-70. <http://dx.doi.org/10.1080/15348458.2013.748430>
- Helsby, G. (1995). Teachers' construction of professionalism in England in the 1990s. *Journal of Education for Teaching*, 21(3), 317-332.
- Hill, M.F. & Evers, G.E. (2016). Moving from student to teacher: Changing perspective about assessment through teacher education. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 57-76). New York: Routledge.
- Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students' conceptions of assessment: Factorial and structural invariance of the SCoA across sex, age, and ethnicity. *European Journal of Psychological Assessment*, 25(1), 30-38.
- Hogan, D. (2011, May-June). *Culture and pedagogy in Singapore: An institutionalist account of the fate of the teach less learn more policy initiative*. Paper presented at the 4th Redesigning Pedagogy: Transforming Teaching, Inspiring Learning International Conference, National Institute of Education, Nanyang Technological University, Singapore.
- Hogan, D., Chan, M., Rahim, R., Maung, A., Chen, L., Zhe, S., & Wenshu, L. (2013). Context and Implications Document for: Assessment And The Logic Of Instructional Practice In Secondary 3 English And Mathematics Classrooms In Singapore. *Review Of Education*, 1(1), 122-123. <http://dx.doi.org/10.1002/rev3.3009>
- Hokka, P., Etelapelto, A. & Rasku-Puttonen, H. (2010). Recent tensions and challenges in teacher education as manifested in curriculum discourse. *Teaching and Teacher Education*, 26(4), 845-853.
- Holt, J. (2005). *The underachieving school*. Boulder, CO: Sentient Publications.
- Hopfenbeck, T., Flórez Petour, M., & Tolo, A. (2015). Balancing tensions in educational policy reforms: large-scale implementation of Assessment for Learning in Norway. *Assessment In Education: Principles, Policy & Practice*, 22(1), 44-60. <http://dx.doi.org/10.1080/0969594x.2014.996524>
- Hornberger, N. & Vaish, V. (2009). Multilingual language policy and school linguistic practice: globalization and English - language teaching in India, Singapore and

- South Africa. *Compare: A Journal of Comparative and International Education*, 39(3), 305-320.
- Huang, S.-C. (2012). Like a bell responding to a striker: Instruction contingent on assessment. *English Teaching: Practice and Critique*, 11(4), 99-119.
- Hui, S. K. F. (2012). Missing conceptions of assessment: Qualitative studies with Hong Kong curriculum leaders. *The Asia-Pacific Education Researcher*, 21(2), 375-383.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923–2931). Oxford: Blackwell.
- Ioannidou, E. (2012). Language policy in Greek Cypriot education: tensions between national and pedagogical values. *Language, Culture and Curriculum*, 25(3), 215-230.
- Irwin, J. W. (1996). *Empowering ourselves and transforming schools: Educators making a difference*. Albany, NY: State University Press.
- James, A. R, Griffin, L. & Dodds, P. (2009). Perceptions of middle school assessment: an ecological view. *Physical Education and Sport Pedagogy*, 14(3), July 2009, 323–334.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Curriculum Journal*, 17(2), 109-138.
- Jamil, M., Topping, K. J. & Tariq, R. H. (2011). Perceptions of University Students regarding Computer assisted assessment. *The Turkish Online Journal of Educational Technology*, 11(3), 267-277.
- Johansson, B., Marton, F., & Svensson, L. (1985). An approach to describing learning as change between qualitatively different conceptions. In L. West & A. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 233–258). Orlando, FL: Academic Press.
- Jones, N., Saville, N., & Salamoura, A. (2016). *Learning Oriented Assessment*. Cambridge: Cambridge Univ. Press.
- Kahl, S. (2005). Where in the World Are Formative Tests? Right Under Your Nose! *Education Week*, 25(4), 11.
- Karp, G. G. & Woods, M. L. (2008). Preservice Teachers' Perceptions About Assessment and Its Implementation. *Journal of Teaching in Physical Education*, 27, 327-346.
- Kaufman, J. H. & Schunn, C.D. (2011). Students' perceptions about peer assessment for writing their origin and impact on revision work. *Instructional Science*, 39, 387–406.
- Kennedy, K. J. (2016). Exploring the influence of culture on assessment: the case of teachers' conception of assessment in Confucian Heritage Cultures. In G. T. L. Brown

& L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 404-419). New York: Routledge.

- Keogh, B. K. (1980). *Advances in Special Education (Volume 2)*. Michigan: Jai Press.
- Khong, L. Y. L & Ng, P. T. (2005). School-Parent Partnerships in Singapore. *Educational Research for Policy and Practice*, 2005(4), 1-11.
- Kim, A., & Kim, H. (2017). The effectiveness of instructor feedback for learning-oriented language assessment: Using an integrated reading-to-write task for English for academic purposes. *Assessing Writing*, 32, 57-71. <http://dx.doi.org/10.1016/j.asw.2016.12.001>
- Kim, M., Tan, A. L. & Talaue, F. T. (2013). New Vision and Challenges in Inquiry-Based Curriculum Change in Singapore. *International Journal of Science Education*, 35(2), 289-311.
- King, C., Patterson, N. & Stolle, E. (2008). Portfolio assessment: Making connections, guiding change. *English Teaching: Practice and Critique*, 7(3), 4-9.
- Klinger, D. A., & Rogers, W. T. (2011). Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing*, 11, 122-143.
- Koh, K. H. & Gong, Wengao. (2008). Alternative assessment and the teaching of Mandarin Chinese to elementary school students in multilingual Singapore. Paper presented at the 2008 American Educational Research Association (AERA) conference in New York, USA.
- Koh, K. H. & Lee, N. L. (2007). The quality of teachers' assessment tasks and student work in the Singapore science classrooms. Paper presented at the 2007 American Educational Research Association (AERA) conference in Chicago Illinois, USA.
- Koh, K. H. & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: an empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy & Practice*, 16(3), 291-318.
- Koh, K. H. & Yeo, J. (2011). *Mastering the art of authentic assessment: from challenges to champions (Volume 2)*. Singapore: Pearson/Prentice Hall.
- Koh, K. H. (2011a). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.
- Koh, K. H. (2011b). *Improving teachers' assessment literacy*. Singapore: Pearson/Prentice Hall.
- Koh, K. H. (2011c). *Mastering the art of authentic assessment: from challenges to champions (Volume 1)*. Singapore: Pearson/Prentice Hall.
- Koh, K. (2012). Teachers' assessment literacy and student learning in Singapore mathematics classrooms. In B. Sriraman, J. Cai, K. Lee, L. Fan, Y. Shimuzu, C.S. Lim,

- & K. Subramanian (Eds.), *The first sourcebook on Asian research in mathematics education: Korea, Singapore, China, Japan, India and Malaysia*. The Montana Mathematics Enthusiast: Monograph Series in Mathematics Education. Charlotte, NC: Information Age Publishing.
- Koh, K. H., Lim, L. & Habib, M. (2010). Building teachers' capacity in classroom-based formative assessment. Paper presented at the 2010 International Association for Educational Assessment (IAEA) conference in Bangkok, Thailand.
- Koh, K., Tan, C & Ng, P.T. (2012). Creating thinking schools through authentic assessment: the case in Singapore. *Educational Assessment, Evaluation and Accountability*, 24, 135-149.
- Korthagen, F. A. J. & Kessels, J. P. A. M. (1999). Linking Theory and Practice: Changing the Pedagogy of Teacher Education. *Educational Researcher*, 28(4), 4-17.
- Kostogriz, A. & Doecke, B. (2013). The ethical practice of teaching literacy: Accountability or responsibility? *Australian Journal of Language and Literacy*, 36(2), 90-98.
- Kuo, E. C. Y. (1996). Confucianism as Political Discourse in Singapore: The Case of an Incomplete Revitalisation Movement. In Wei-Ming. Tu (Ed.), *Confucian Traditions in East Asian Modernity: Moral Education and Economic Culture in Japan and the Four Mini-Dragons* (pp. 294-309). Cambridge, Massachusetts: Harvard University Press.
- Krathwohl, D.R., Bloom, B.S., & Masia, B.B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain*. New York: David McKay Co.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*: Sage Publications.
- Kwon, S., Lee, M., & Shin, D. (2015). Educational assessment in the Republic of Korea: lights and shadows of high-stake exam-based education system. *Assessment In Education: Principles, Policy & Practice*, 24(1), 60-77. <http://dx.doi.org/10.1080/0969594x.2015.1074540>
- Lakoff, G. (1993). The contemporary theory of metaphor. *Metaphor and Thought*. A. Ortony. Cambridge, Cambridge University Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York, NY: Basic Books.
- Lantolf, J. & Poehner, M. (2004). Dynamic assessment of L2 development: bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49-72.
- Larsson, J., & Holmström, I. (2007). Phenomenographic or phenomenological analysis: does it matter? Examples from a study on anaesthesiologists' work. *International Journal Of Qualitative Studies On Health And Well-Being*, 2(1), 55-64. doi: 10.1080/17482620601068105

- Larsson, J., Holmström, I., & Rosenqvist, U. (2003). Professional artist, good Samaritan, servant and co-ordinator: four ways of understanding the anaesthetist's work. *Acta Anaesthesiologica Scandinavica*, 47, 787-793.
- Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education*, 37(2), 137-147.
- Lecos, M. A., Cassella, C., Evans, C., Leahy, C., Liess, E., & Lucas, T., (2000). Empowering Teacher Leadership in Professional Development Schools. *Teaching and Change*, 8(1), 98-113.
- Lee, A. N., & Nie, Y. (2013). Understanding teacher empowerment: Teachers' perceptions of principal's and immediate supervisor's empowering behaviours, psychological empowerment and work-related outcomes. *Teaching and Teacher Education*, 41, 67-79.
- Lee, H. L. (2004). Our future of opportunity and promise. Singapore Government Press Release. Address by Prime Minister Lee Hsien Loong at the 2004 National Day Rally at the University Cultural Centre, National University of Singapore, 22 August.
- Lee, Y. (2009). Not if but when pedagogy collides with culture in Singapore. *Pedagogies: An International Journal*, 5(1), 17-26. <http://dx.doi.org/10.1080/15544800903406274>
- Lefstein, A. (2008). Changing Classroom Practice Through the English National Literacy Strategy: A Micro-Interactional Perspective. *American Educational Research Journal*, 45(3), 701-737. <http://dx.doi.org/10.3102/0002831208316256>
- Leong, W. S., Cheng, Y. S., & Tan, K. (2014). Setting the context assessment and learning in schools. In W.S. Leong, Y. S. Cheng, & K., Tan (Eds.), *Assessment and Learning in Schools* (pp. 1-6). Singapore: Pearson.
- Leong, W. S., & Tan, K. (2014). What (more) can, and should, assessment do for learning? Observations from 'successful learning context' in Singapore. *The Curriculum Journal*, 25(4), 593-619.
- Li, W.S., & Hui, S.K.F. (2007). Conceptions of assessment of Mainland China college lecturers: A technical paper analysing the Chinese version of COA-III. *The Asia-Pacific Education Researcher*, 16(2), 185-198.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.
- Lindgren, C. L. (1959). Learning Theory and Teaching Practice. *Educational Leadership*, 16(6), 333-336.
- Lindsay, G., & Lewis, A. (2003). An Evaluation of the Use of Accredited Baseline Assessment Schemes in England. *British Educational Research Journal*, 29(2), 149-167. <http://dx.doi.org/10.1080/0141192032000060902>

- Linn, R. (2000) Assessments and accountability, *Educational Researcher*, 29(2), 4–16.
- Little, D. 1995. Learning as dialogue: The dependence of learner autonomy on teacher autonomy. *System*, 23(2), 175–182.
- Liu, Chaoying. (2011). Dui xin jia po hua wen ke cheng ping jia fang shi de ji dian kan fa (Views on the modes of assessment of the Chinese Language curriculum in Singapore). *Hua wen lao shi (Journal of Chinese Language Teaching in Singapore)*, 55(24), 18-23.
- Liu, S. N., & Feng, D. M. (2015). How culture matters in educational borrowing? Chinese teachers' dilemmas in a global era. *Cogent Education*, 2(1).
- Liu, P. D., Zhou, Y. & McBride-Chang, C. (2010). Morphological awareness. In Jianguo Chen, Chuang Wang & Jinfa Cai (eds.), *Teaching and learning Chinese: issue and perspectives* (pp.237-256). New Carolina: Information Age Publishing.
- Liu, S. (2013). Pedagogical Content Knowledge: A case study of ESL teacher educator. *English Language Teaching*, 6(7), 128-138.
- Luk-Fong, P. Y. Y., & Brennan, M. (2010). Teachers' experience of secondary education reform in Hong Kong. *International Journal of Educational Reform*, 19(2), 128-153.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: what do they really mean to the raters? *Language Testing*, 19(3) 246–276.
- MacBeath, J., & Galton, M. (2004). *A Life in Teaching: Finding time for learning*. London: NUT.
- McDonald, B., & Boud, D. (2003). The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations. *Assessment In Education: Principles, Policy & Practice*, 10(2), 209-220.
- McKay, P., & Brindley, G. (2007). Educational Reform and ESL Assessment in Australia: New Roles and New Tensions. *Language Assessment Quarterly*, 4(1), 69-84.
- Macdonald, B., & Walker, R. (1977). Case-study and the social philosophy of educational research. In D. Hamilton et al. (Eds), *Beyond the Numbers Game: a Reader in Educational Evaluation* (pp. 181-189). Basingstoke: Macmillan.
- Maclellan, E. (2001). Assessment for Learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318.
- Maeroff, G. I. (1989). The principles of teacher empowerment. *The Education Digest*, 54(6), 6-9.
- Mak, S. H. (2011). Tensions Between Conflicting Beliefs of an EFL Teacher in Teaching Practice. *RELC Journal*, 42(1), 53-67.
- Mansell, W., James, M. & the Assessment Reform Group (2009). *Assessment in schools. Fit for purpose?* A Commentary by the Teaching and Learning Research Programme.

London: Economic and Social Research Council, Teaching and Learning Research Programme.

Martin, E., Prosser, M., Trigwell, K., Lueckenhausen, G., & Ramsden, P. (2001). Using Phenomenography and Metaphor to Explore Academics' Understanding of Subject Matter and Teaching. In C Rust (Ed). *Improving Student Learning*. Oxford: The Oxford centre for Staff and Learning Development, 325-336.

Mason J. (2002) *Qualitative Researching*. London: SAGE.

Massell, D., Kirst, M., & Hoppe, M. (1997). *Persistence and change: Standards-based reform in nine states*. New Brunswick, NJ: Consortium for Policy Research in Education.

Marguerite, G. L., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in Educational Research: From Theory to Practice*. San Francisco, Calif. : Jossey-Bass.

Marton, F. (1981). Phenomenography-Describing conceptions of the world around us. *Instructional Science*, 10(2), 177-200.

Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought*, 21(3), 28-43.

Marton, F. (1994). Phenomenography In T. Husun & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., Vol. 8, pp. 4424-4429). Oxford, U.K.: Pergamon.

Marton, F. (2000). The structure of awareness. In J. Bowden & E. Walsh (Eds.), *Phenomenography* (pp.102-116). Melbourne: RMIT Publishing.

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah, New Jersey: Lawrence Erlbaum.

Marton, F., & Pong, W. (2005). On the unit of description in phenomenography. *Higher Education Research & Development*, 24(4), 335-348. <http://dx.doi.org/10.1080/07294360500284706>

Marzano, R. J., Pickering, D. J., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Alexandria, VA: ASCD.

McKay, P. & Brindley, G. (2007). Educational Reform and ESL Assessment in Australia: New Roles and New Tensions. *Language Assessment Quarterly*, 4(1), 69-84.

McMillan, J. H. & Nash, S. (2000). Teacher Classroom Assessment and Grading Practices Decision Making. Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April25-27, 2000).

McMillan, J. H. & Schumacher, S. (2006). *Research in Education: Evidence-based enquiry*. Boston, MA: Pearson Education.

- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McNamara, T. (1997). Performance testing. In Clapham, C. Corson, D. (Eds.), *Encyclopedia of language and education: Volume 7 Language testing and assessment* (pp. 131-139). Dordrecht, NL: Kluwer Academic.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Miller, P. (2008). Validity. In L. M. Given (ed.), *The Sage encyclopedia of qualitative research methods (Volume 1&2)*, pp.909-910. Thousand Oaks, CA: SAGE Publications, Inc.
- Ministry of Education. (2009). *Report of the Primary Education Review and Implementation PERI Committee*. Singapore. Primary Education Review and Implementation Committee.
- Ministry of Education. (2010). *2010 Mother Tongue Languages Review Committee Report*.
- Ministry of Education. (2010). MOE to Enhance Learning of 21st Century Competencies and Strengthen Art, Music and Physical Education. [Press Release], Retrieved March 09, 2016, from <http://www.moe.gov.sg/media/press/2010/03/moe-to-enhance-learning-of-21s.php>
- Ministry of Education. (2018). 'Learn For Life' – Preparing Our Students To Excel Beyond Exam Results. [Press Release], Retrieved November 02, 2018, from <https://www.moe.gov.sg/news/press-releases/-learn-for-life---preparing-our-students-to-excel-beyond-exam-results>
- Ministry of Education-OPAL. (2018). Assessment concepts in assessment portal. Retrieved September 20, 2018, from MOE Intranet website. <https://subjects.opal.moe.edu.sg/cos/o.x?c=/subjects/pagetree&func=view&rid=9078>
- Mok, J. (2011). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal*, 65(3), 230-239.
- Morgan, D. (1997). *Focus groups as qualitative research*. Thousand Oaks, CA:Sage.
- Moss, C. & Brookhart, S. (2009). *Advancing formative assessment in every classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.

- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109–162.
- Murphy, M.S., Pinnegar, E. & Pinnegar, S. (2011). Exploring Ethical Tensions on the Path to Becoming a Teacher. *Teacher Education Quarterly*, 38(4), 97-113.
- National Institute of Education (NIE). (2012). *A Teacher Education Model for the 21st Century (TE21): NIE's journey from concept to realisation-An Implementation Report*. Singapore: NIE.
- National Education Association. (2003). *Balanced assessment: The key to accountability and improved student learning*. Washington, DC: National Education Association.
- Newton, D. & Newton, L. (2009). Some student teachers' conceptions of creativity in school science. *Research in Science and Technological Education*, 27(1), 45-60.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170.
- Ng, P. T. (2008). Educational reform in Singapore: from quantity to quality. *Educational Research for Policy and Practice*, 7(1), 5-15.
- Nguyen, P. M., Terlouw, C., & Pilot, A. (2006). Culturally appropriate pedagogy: The case of group learning in a Confucian heritage culture context. *Intercultural Education*, 17(1), 1–19.
- Nicola, C., Novakowskia, J., Ghaleba, F. & Beirstob, S. (2010). Interweaving Pedagogies of Care and Inquiry: Tensions, dilemmas and possibilities. *Studying Teacher Education*, 6(3), 235–244.
- Norashikin, B. M. N., Muniandy, M. K., Shanmugan, S. K. K & Mathai, E. J. (2010). Upper Primary Teachers' Perceptions of PSLE English Oral Assessment. *English Language Teaching*, 3(4), 142-151.
- Noss, R., Goldstein, H., & Hoyles, C. (1989). Graded Assessment and Learning Hierarchies in Mathematics. *British Educational Research Journal*, 15(2), 109-120.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University.
- O'Dwyer, S. (2016). Deflating the 'Confucian Heritage Culture' thesis in intercultural and academic English education. *Language, Culture And Curriculum*, 30(2), 198-211. <http://dx.doi.org/10.1080/07908318.2016.1259321>
- OECD (2011). *Preparing Teachers and Developing School Leaders for 21st Century - Lessons from around the world (Background Report for the International Summit on the Teaching Profession)*. PISA: OECD Publishing.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380.

- Padgett, D. K. (2008). Advocacy research. In L. M. Given (ed.), *The Sage encyclopedia of qualitative research methods (Volume 1&2)* (pp.8-11). Thousand Oaks, CA: SAGE Publications, Inc.
- Paechter, C. (1995). Doing the Best for the Students: dilemmas and decisions in carrying out statutory assessment tasks. *Assessment in Education: Principles, Policy & Practice*, 2(1), 39-52.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332.
- Pang, M.F. (2003). Two Faces of Variation: on continuity in the phenomenographic movement. *Scandinavian Journal of Educational Research*, 47(2), 145-156.
- Pang, M.F. (2008). Xian xiang tu xi xue yu bian yi li lun (Phenomenography and Variation Theory). *Jiao yu xue bao (Journal of Educational Studies)*, 4(5), 33-38.
- Parke, C. S. & Lane, S. (2007). Students' perceptions of a Maryland state performance assessment. *The Elementary School Journal*, 107(3), 305-324.
- Patton, C. (2011). 'Some kind of weird, evil experiment': student perceptions of peer assessment. *Assessment & Evaluation in Higher Education*, 37(6), 719-731.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: the science and design of educational assessment*. Washington: National Academy Press.
- Peterson, E. & Irving, E. (2007). *Conceptions of Assessment and Feedback*. Auckland UniServices Limited.
- Petrarca, D., & Bullock, S. M. (2014). Tensions between theory and practice: interrogating our pedagogy through collaborative self-study. *Professional Development in Education*, 40(2), 265-281.
- Phipps, S. & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37(3), 380-390.
- Pillen, M., Beijaard, D. & den Brok, P. (2013). Professional identity tensions of beginning teachers. *Teachers and Teaching: theory and practice*, 19(6), 660-678.
- Pinto, L. (2016). Tensions and fissures: the politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, 27(1), 95-112. <http://dx.doi.org/10.1080/09585176.2016.1140061>
- Pitcher, R., & Åkerlind, G. (2009). Post - Doctoral Researchers' Conceptions of Research: A Metaphor Analysis. *International Journal For Researcher Development*, 1(2), 160-172. doi: 10.1108/1759751x201100009
- Poehner, M. (2008). *Dynamic assessment*. New York: Springer.

- Poehner, E. M. & Infante, P. (2017). Dynamic assessment in the language classroom. In: Dina Tsagari and Jayanti Banerjee (eds.), *Hand book of Second Language Assessment*. Berlin: DeGruyter Mouton.
- Poole, A. (2016). 'Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education*, 3(1). <http://dx.doi.org/10.1080/2331186x.2016.1156242>
- Popham, W. J. (1995). *Classroom assessment: what teachers need to know*. USA: Allyn and Bacon.
- Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48, 4–11.
- Pratt, D. (1992). Conceptions of Teaching. *Adult Education Quarterly*, 42(4), 203-220. <http://dx.doi.org/10.1177/074171369204200401>
- Pratt, N. (2016). Neoliberalism and the (internal) marketisation of primary school assessment in England. *British Educational Research Journal*, 42(5), 890-905. <http://dx.doi.org/10.1002/berj.3233>
- Prosser, M. (2000). Using phenomenographic research methodology in the context of research in teaching and learning. In J. A. Bowden & E. Walsh (Eds.), *Phenomenography* (pp. 34-46). Melbourne: RMIT University Press.
- Prosser, M., Martin, E., Trigwell, K., Ramsden, P., & Lueckenhausen, G. (2005). Academics' experiences of understanding of their subject matter and the relationship of this to their experiences of teaching and learning. *Instructional Science*, 33(2), 137-157. doi: 10.1007/s11251-004-7687-x.
- Qi, Yong-hua, Xie, Xijin & Cen, Shaoji. (2005). *Bian yi li lun yu xue xi kong jian (The theory of variations and outcome space)*. Xiang gang: Xiang gang da xue chu ban she.
- Randall & Zundel, (2012). Students' Perceptions of the Effectiveness of Assessment Feedback as a Learning Tool in an Introductory Problem-solving Course. *The Canadian Journal for the Scholarship of Teaching and Learning*, 3(1), <http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.3>
- Ratnam-Lim, C., & Tan, K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment In Education: Principles, Policy & Practice*, 22(1), 61-78. doi: 10.1080/0969594x.2014.1001319
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343-354. <http://dx.doi.org/10.1080/0141192990250305>

- Remesal, A. (2007). Educational Reform and Primary and Secondary Teachers' Conceptions of Assessment: The Spanish Instance, Building upon Black and Wiliam. *Curriculum Journal*, 18(1), 27-38.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482.
- Restorff, D., Sharpe, M., Abery, B., Rodriguez, M. & Kim, N. K. (2012). Teacher perceptions of alternate assessments based on alternate achievement standards: Results from a three-state survey. *Research and Practice for Persons with Severe Disabilities*, 37(3), 185-198.
- Richards, J.C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. New York: Cambridge Univ. Press.
- Rothman, R. (2010). *Principles for a comprehensive assessment system*. Washington, DC: Alliance for Excellent Education.
- Rowntree, D. (1987). *Assessing Students: How shall we know them?* London: Kogan Page.
- Rowold, J. (2008). Instrument Development for Esthetic Perception Assessment. *Journal of Media Psychology: Theories, Methods, and Applications*, 20(1), 35-40.
- Rubdy, R. & Tupas, T. R. F. (2009). Research in applied linguistics and language teaching and learning in Singapore (2000–2007). *Language Teaching*, 42 (3), 317-340.
- Ruiz-Primo, M.A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation*. CSE Report 677. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Saban, A., Kocbeker, B. N., & Saban, A. (2007). Prospective teachers' conceptions of teaching and learning revealed through metaphor analysis. *Learning and Instruction*, 17(2), 123–139.
- Sadeghi, K., & Rahmati, T. (2017). Integrating assessment as, for, and of learning in a large-scale exam preparation course. *Assessing Writing*, 34, 50-61. <http://dx.doi.org/10.1016/j.asw.2017.09.003>
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <http://dx.doi.org/10.1007/bf00117714>
- Saito, H., & Inoi, S. (2017). Junior and Senior High School EFL Teachers' Use of Formative Assessment: A Mixed-Methods Study. *Language Assessment Quarterly*, 14(3), 213-233. <http://dx.doi.org/10.1080/15434303.2017.1351975>
- Säljö, R. (1996). Minding action-Conceiving of the world versus participating in cultural practices. In G.Dall'Alba & B. Hasselgren (Eds.), *Reflections on Phenomenography: Toward a methodology?* (Vol. 109, pp. 19-33. Gothenburg: Acta Universitatis Gothoburgensis.

- Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: SAGE Publications.
- Sandbergh, J. (1997). Are Phenomenographic Results Reliable?. *Higher Education Research & Development*, 16 (2), 203-212.
- Sandholtz, J. h. (2011). Preservice teachers' conceptions of effective and ineffective teaching practices. *Teacher Education Quarterly*. 38(3), 27.
- Sato, M., Wei, R.C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669–700.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327.
- Scriven, M. (1967). The Methodology of Evaluation. In Ralph W. Tyler et al., (eds.), *Perspectives in Evaluation* (pp. 39-83). American Educational Research Association Monograph Series on Curriculum Evaluation, No. 1. Chicago, Ill.: Rand McNally.
- Seawright, J. & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly*, 61(2), 294-308.
- Shapira-Lishchinsky, O. (2011). Teachers' critical incidents: Ethical dilemmas in teaching practice. *Teaching and Teacher Education*, 27 (3), 648-656.
- Sharpe, L. & Gopinathan, S. (2002). After effectiveness: new directions in the Singapore school system?, *Journal of Education Policy*, 17(2), 151-166.
- Shavelson, R., Young, D., Ayala, C., Brandon, P., Furtak, E., & Ruiz-Primo, M. et al. (2008). On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers. *Applied Measurement In Education*, 21(4), 295-314.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C.A. Dwyer (ed.), *The future of assessment: Shaping teaching and learning* (pp. 279–303). New York: Erlbaum.
- Shihiba, S, E, S. (2011). *An Investigation of Libyan EFL Teachers' Conceptions of the Communicative Learner-Centred Approach in Relation to their Implementation of an English Language Curriculum Innovation in Secondary Schools*. Doctoral thesis, Durham University.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the use of language tests*. Harlow, UK: Longman.

- Short, P. M. (1994). Defining teacher empowerment. *Education*, 114(4), 488-492.
- Short, P. M., & Rinehart, J. S. (1992). School participant empowerment scale: assessment of level of empowerment within the school environment. *Journal of Educational and Psychological Measurement*, 52(4), 951-960.
- Silver, R. E. (2005). The discourse of linguistic capital: Language and economic policy planning in Singapore. *Language Policy*, 4, 47-66.
- Simons, H. (1996) The paradox of case study. *Cambridge Journal of Education*, 26(2), 225-40.
- Simpson, M. & Grudnoff, L. (2013). From preparation to practice: Tensions and connections. *Waikato Journal of Education*, 18(1), 71-82.
- Sin, S. (2010). Considerations of Quality in Phenomenographic Research. *International Journal of Qualitative Methods*, 9(4), 305-319. [http://dx.doi.org/10.1177/160940691000900401`](http://dx.doi.org/10.1177/160940691000900401)
- Sinclair, B., McGrath, I., & Lamb, T. (2001). *Learner autonomy, teacher autonomy*. Harlow: Longman.
- Singapore Centre for Chinese Language (SCCL). (2010). *Executive Director's Message*. Retrieved from <http://en.sccl.sg/cos/o.x?c=/wbn/pagetree&func=view&rid=1108582>
- Skehan, P. (1998a). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268-286.
- Slawter, L. (2004). *Students' Perceptions of Assessment in the Elementary School*. EdD Dissertation.
- Smith, K. (2016). Cooperative Learning About Assessment for Learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 181-197). Switzerland: Springer, Cham.
- Song, E. & Koh, K. H. (2010). Assessment for Learning: understanding teachers' beliefs and practices. Paper presented at the 2010 International Association for Educational Assessment (IAEA) conference in Bangkok, Thailand.
- Spolsky, B. (1978). Introduction: Linguistics and language testers. In B. Spolsky (Ed.), *Advances in language testing series: 2*. Arlington, VA: Center for Applied Linguistics.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE.
- Starr, D. (2009). Chinese Language Education in Europe: the Confucius Institutes. *European Journal of Education*, 44(1), Part 1, 65-82.
- Stemhagen, K. (2011). Democracy and School Math: teacher belief-practice tensions and the problem of empirical research on educational aims. *Democracy & Education*, 19 (2), 1-13.

- Stern, E. (2013). Editorial. *Evaluation*, 19(3), 213-216.
- Stiggins, R. J. (1991). Assessment Literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. J. (1997). *Student-Centered Classroom Assessment*. Columbus: Prentice Hall, Inc.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stillman, J. & Anderson, L. (2011). To Follow, Reject, or Flip the Script: Managing Instructional Tension in an Era of High-Stakes Accountability. *Language Arts*, 89(1), 22-37.
- Stobart, G. (2008) *Testing Times, The Uses and Abuses of Assessment*. Abingdon, Routledge.
- Stobart, G. (2010). Assessment fit-for-the-future. Paper presented at the IAEA Annual Conference 2010, Bangkok, Thailand.
- Stobart, G. and Gipps, C. (1997) *Assessment: A Teacher's Guide to the Issues*. London: Hodder and Stoughton.
- Stoyonoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. Alexandria, VA: TESOL.
- Strauss, P. & U, A. (2007). Group assessments: dilemmas facing lecturers in multicultural tertiary classrooms. *Higher Education Research & Development*, 26(2), 147-161.
- Stroud, C. & Wee, L. (2007). Consuming identities: Language planning and policy in Singaporean late modernity. *Lang Policy*, 6(2), 253-279. <http://dx.doi.org/10.1007/s10993-007-9047-6>
- Sturman, A. (1994) Case study methods. In J. P. Keeves (ed.), *Educational Research, Methodology, and Measurement: an International Handbook, 2nd edn.* (pp. 61-66). Oxford: Pergamon.
- Suurtamm, C., & Koch, M. (2014). Navigating dilemmas in transforming assessment practices: experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation And Accountability*, 26(3), 263-287. <http://dx.doi.org/10.1007/s11092-014-9195-0>
- Svensson, L. (1997). Theoretical Foundations of Phenomenography. *Higher Education Research & Development*, 16(2), 159-171.
- Sweetland, S. R., & Hoy, W. K. (2000). School characteristics and educational outcomes: toward an organizational model of student achievement in middle schools. *Educational Administration Quarterly*, 36(5), 703-729.

- Talanquer, V., Tomanek, D., & Novodvorsky, I. (2007). Revealing student teachers' thinking through dilemma analysis. *Journal of Science Teacher Education*, 18(3), 399-421.
- Tan, A. & Leong, W. F. (2014). Mapping Curriculum Innovation in STEM Schools to Assessment Requirements: Tensions and Dilemmas. *Theory Into Practice*, 53(1), 11-17.
- Tan, C. (2006). Change and Continuity: Chinese Language Policy in Singapore. *Language Policy*, 5(1), 41-62.
- Tan, C. (2008). Globalisation, the Singapore state and educational reforms: towards performativity. *Education, Knowledge and Economy*, 2(2), 111-120.
- Tan, C. (2012). 'Our shared values' in Singapore: A Confucian perspective. *Educational Theory*, 62(4), 449-463. doi: 10.1111/j.1741-5446.2012.00456.x
- Tan, C. (2017). Confucianism and education. In Noblit, G. (Ed.), *Oxford Research Encyclopedia of Education* (pp. 1-18). New York: Oxford University Press.
- Tan, K. H. K. (2004). Does student self-assessment empower or discipline students? *Assessment and Evaluation in Higher Education*, 29 (6), 651-662.
- Tan, K. H. K. (2007a). *Alternative assessment in schools: a qualitative approach*. Singapore: Pearson/Prentice Hall.
- Tan, K. H. K. (2007b). Conceptions of self-assessment: What is needed for long term learning? In D. Boud & N. Falchikov (Eds.), *Rethinking assessment for higher education: Learning for the longer term* (pp. 114-127). London: Routledge.
- Tan, K. H. K. (2008a). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research & Development*, 27(1), 15-29.
- Tan, K. H. K. (2008b). Rethinking TLLM and its consequential effects on assessment. In J. Tan & P. T. Ng (Eds.), *Thinking schools, learning nation: Contemporary issues and challenges* (pp. 246-257). Singapore: Pearson Education South Asia Pte Ltd.
- Tan, K. H. K. (2009). Variation theory and the different ways of experiencing educational policy. *Educational Research for Policy and Practice*, 8(2), 95-109.
- Tan, K. H. K. (2010). *Reframing alternative assessment in schools: a research based perspective*. Singapore: Prentice Hall.
- Tan, K. H. K. (2011a). Assessment for learning in Singapore—Unpacking its meanings and identifying some areas for improvement. *Educational Research for Policy and Practice*, 10(2), 91–103.
- Tan, K. H. K. (2011b). Assessment reform in Singapore—Enduring, sustainable or threshold?. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp.75–88). London: Springer.

- Tan, K. H. K. (2012). How teachers understand and use power in alternative assessment. *Education Research International*, Vol. 2012. <http://www.hindawi.com/journals/edu/2012/382465/>
- Tan, K. H. K. (2013). Variation in teachers' conceptions of alternative assessment in Singapore primary schools. *Educational Research for Policy and Practice*, 12(1), 21-41.
- Tan, K. H. K., & Koh, K. H. (2008). *Authentic assessment in schools*. Singapore: Pearson/Prentice Hall.
- Tan, K. H. K. & Prosser, M. (2004). Qualitatively different ways of differentiating student achievement: A phenomenographic study of grade descriptors. *Assessment and Evaluation in Higher Education*, 29(3), 267-282.
- Tan, O. S. (2007). *Teach less, learn more (TLLM) school-based curriculum innovation : research reports*. Singapore: Co-published by Curriculum Policy and Pedagogy Unit, Curriculum Planning and Development Division, Ministry of Education [and] Educational Research Association of Singapore.
- Tan, O. S. (2012). *Teacher education frontiers: international perspectives on policy and practice for building new teacher competencies*. Singapore: Cengage Learning.
- testing: research contexts and methods (pp. 129–146). Mahwah, NJ: Erlbaum.
- Taras, M. (2005). ASSESSMENT – Summative and formative – Some theoretical reflections. *British Journal Of Educational Studies*, 53(4), 466-478. <http://dx.doi.org/10.1111/j.1467-8527.2005.00307.x>
- Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of Further and Higher Education*, 33(1), 57-69. <http://dx.doi.org/10.1080/03098770802638671>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412.
- Taylor, J., & Wallace, G. (1990). Some Dilemmas in Implementing the Criteria for Continuous Assessment in GCSE English. *British Journal Of Sociology Of Education*, 11(1), 49-64. <http://dx.doi.org/10.1080/0142569900110104>
- Thanh Pham, T. & Renshaw, P. (2014). Formative assessment in Confucian heritage culture classrooms: activity theory analysis of tensions, contradictions and hybrid practices. *Assessment & Evaluation In Higher Education*, 40(1), 45-59. <http://dx.doi.org/10.1080/02602938.2014.886325>
- Tharman, S. (2004). *To light a fire: Enabling teachers, nurturing students*. Speech by Mr Tharman Shanmugaratnam, Minister for Education, at the MOE Work Plan Seminar 2004 at the Ngee Ann Polytechnic Convention Centre, Singapore, 29 September. <http://www3.moe.edu.sg/pld/ine/ine.htm>.

- Thomas, G. (2011). A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. *Qualitative Inquiry*, 17(6) 511–521.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York: MacMillan.
- Thompson, M., & Wiliam, D. (2008). Tight but loose: a conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: scaling up teacher professional development in diverse contexts* (RR-08-29, pp. 1-44). Princeton, NJ: Educational Testing Service.
- Tillema, H., & Kremer-Hayon, L. (2005). Facing dilemmas: teacher-educators' ways of constructing a pedagogy of teacher education. *Teaching in Higher Education*, 10(2), 203-217.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment*. Maidenhead, Philadelphia: Open University Press.
- Tsui, A. (2004). The shared space of learning. In F. Marton & A. Tsui (Eds.), *Classroom discourse and the space of learning* (pp. 165-186). New Jersey: Lawrence Erlbaum Associates.
- Triggs, P., & Pollard, A. (2000). *What Pupils Say: Changing Policy and Practice in Primary Education*. London: Continuum International Publishing Group Ltd.
- Trigwell, K. (2000). A phenomenographic interview on phenomenography. In J. A. Bowden & E. Walsh (Eds.), *Phenomenography* (pp. 62-82). Melbourne: RMIT University Press.
- Trilling, B., & Fadel, C. (2009). *21st century skills: learning for life in our times*. San Francisco, Calif.: Jossey-Bass.
- Troman, G. (1989). Testing Tensions: the politics of educational assessment. *British Educational Research Journal*, 15(3), 279-295.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment and Evaluation in Higher Education*, 31(5), 501-521.
- Tunstall, P., & Gipps, C. (1996). 'How does your teacher help you to make your work better?' Children's understanding of formative assessment. *Curriculum Journal*, 7(2), 185-203.
- Turner, J. (2011). *Language in the Academy*. Bristol: Multilingual Matters.
- Vandeyar, S. (2005). Conflicting Demands: Assessment Practices in Three South African Primary Schools Undergoing Desegregation. *Curriculum Inquiry*, 35(4), 461-481.
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, 34(2), 239-255.

- Walsh, E. (2000). Phenomenographic analysis of interview transcripts. In J. Bowden & E. Walsh (Eds.), *Phenomenography* (pp.19-33). Melbourne: RMIT Publishing.
- Wang, Jing-Ru, Kao, Huey-Lien & Lin, Sheau-Wen. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education*, 26(3), 522-529.
- Wang, S. C. (2004). *Bi-literacy resource eco-System of intergenerational transmission of heritage language and culture: An ethnographic study of a Chinese community in the United States*. Unpublished Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA.
- Wang, S. C. (2007). Building societal capital: Chinese in the US. *Language Policy*, 6(1), 27–52.
- Wang, Xueli & Hurley, S. (2012). Assessment as a Scholarly Activity? Faculty Perceptions of and Willingness to Engage in Student Learning Assessment, *The Journal of General Education*, 61(1), 1-15.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: research contexts and methods* (pp. 129–146). Mahwah, NJ: Erlbaum.
- Watkins, D.A. & Biggs, J.B. (Eds.) (1996). *The Chinese Learner: Cultural, psychological and contextual influences*. Hong Kong: Comparative Education Research Centre and Victoria, Australia: The Australian Council for the Educational Research.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written response. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.
- Webb, M. & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 16(2), 165-184.
- Wee, L. (2003). Linguistic Instrumentalism in Singapore. *Journal of Multilingual and Multicultural Development*, 24(3), 211-224.
- Wegner, E. & Nückles, M. (2015). Training the brain or tending a garden? Students' metaphors of learning predict self-reported learning patterns. *Frontline Learning Research*, 3(4), 95-109.
- Westbury, I. (2008). The making of formal curricula: Why do states make curricula, and how? In Connelly, F. M., He, M. F., & Phillion, J. (Eds.), *Handbook of curriculum and instruction* (pp. 45-65). Thousand Oaks, CA: Sage.
- White, E. (2009). Student Perspectives of Peer Assessment for Learning in a Public Speaking Course. *Asian EFL Journal*, 33, http://www.asian-efl-journal.com/pta_January_09.pdf
- Wiggins, G. (1989). Teaching to the Authentic Task. *Educational Leadership*, 46(7), 41-47.

- Wiggins, G. (2012). 7 keys to effective feedback. *Educational Leadership*, 70(1), 10-16.
- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment In Education: Principles, Policy & Practice*, 11(1), 49-65.
- William, D. & Thompson, M. (2007). Integrating assessment with learning: what will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). New Jersey: Lawrence Erlbaum Associates.
- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: an analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72(2), 131-175.
- Wininger, S. (2005). Using Your Tests to Teach: Formative Summative Assessment. *Teaching Of Psychology*, 32(3), 164-166. http://dx.doi.org/10.1207/s15328023top3203_7
- Winterbottom, M., Brindley, S., Taber, K.S., Fisher, L.G., Finney, J., & Riga, F. (2008). *Understanding trainee teachers' developing values and practice in relation to assessment*. Paper presented at the European Conference on Educational Research. Gothenburg, Sweden. 10th-12th September 2008.
- Wolf, D., Bixby, J., Glenn, H., & Gardner, H. (1991). To use their minds well: investigation new forms of students' assessment. *Review of Research in Education*, 17, 31-74.
- Wong, H. M. (2012). *The perceptions of Singaporean teachers and students towards academic self-assessment*. PhD Dissertation, NTU, NIE.
- Wu, Ying-cheng (Goh, Yeng Seng). (2009). Duo yu huan jing xia han yu zuo wei di er yu wen jiao xue xin jia po jin yan yu mo shi (TCSL in a multi-lingual situation: a case study of TCSL in Singapore). *Yun nan shi fan da xue xue bao dui wai han yu jiao xue yan jiu ban (Journal of Yunnan Normal University, TCSL Teaching and Research Version)*, 2009(2), 6-12.
- Wylie, E., Goe, L., Leusner, D., Lyon, C., Tomic, C., & Wylie, E. et al. (2008). TIGHT BUT LOOSE: SCALING UP TEACHER PROFESSIONAL DEVELOPMENT IN DIVERSE CONTEXTS. *ETS Research Report Series*, 2008(1), i-141. doi: 10.1002/j.2333-8504.2008.tb02115.x
- Xiao Yun. (2011). Chinese language in the United States: an ethnolinguistic perspective. In Linda Tsung & Ken Cruickshank (Eds.), *Teaching and learning Chinese in global contexts: multimodality and literacy in the New Media Age* (pp.181-195). London: Continuum.
- Xiao, Y. (2017). Formative Assessment in a Test-Dominated Context: How Test Practice Can Become More Productive. *Language Assessment Quarterly*, 14(4), 295-311. <http://dx.doi.org/10.1080/15434303.2017.1347789>

- Xiong, H., Li, L., & Qu, Y. (2015). Exploring EFL Teachers' Cognitive Models Through Metaphor Analysis. *SAGE Open*, 5(4), 1-11. <http://dx.doi.org/10.1177/2158244015608900>
- Xing, F. Y. (1993). *Xian dai han yu (Modern Chinese)*. Beijing: Gao deng jiao yu chu ban she.
- Xu, Feng. (2008) Xin jia po hua yu di wei de bian qian yu hua wen jiao yu suo mian lin de tiao zhan (Learning from the changes to the Chinese Language status and challenges faced by the Chinese Language education). *Zhong wen zhi xue zhi dao (Review and Research on Chinese Literature)*, 2008(6), 13-19.
- Yan, C. (2012). 'We can only change in a small way': A study of secondary English teachers' implementation of curriculum reform in China. *Journal of Educational Change*, 13(4), 431-447.
- Yang, Xia. (2008). Xin jia po zhong xue hua wen ke shi jiao xue shi shi de te dian ji qi shi (The features of Singapore secondary school Chinese Language classroom implementation). *Xian dai yu wen jiao xue yan jiu (Journal of Modern Language Teaching Research)*, 2008(11), 10-11.
- Yates, L. & Holt, B. (2009). "Under pressure I fall back to being a teacher..." Confronting Contending Desires for Schooling and Teaching in a Middle School Reform Project. *The Australian Educational Researcher*, 36(1), 27-42.
- Yin, R. K. (2009). *Case study research: Design and methods (4th ed.)*. Thousand Oaks, CA: SAGE.
- Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, 19(3), 321-332.
- Yin, X., & Buck, G. A. (2015). There is another choice: an exploration of integrating formative assessment in a Chinese high school chemistry classroom through collaborative action research. *Cultural Studies of Science Education*, 10(3), 753-762.
- Yu, N. (2003). Chinese metaphors of thinking. *Cognitive Linguistics*, 14(2-3), 141-165. doi: 10.1515/cogl.2003.006
- Zembylas, M., & Papanastasiou, E. C. (2005). Modeling teacher empowerment: the role of job satisfaction. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 11(5), 433-459.
- Zgaga, P. (2013). The future of European teacher education in the heavy seas of higher education. *Teacher Development*, 17(3), 347-361.
- Zhang, Yanyan. (2010). A Narrative Analysis of Teachers' Beliefs about Chinese Language Curriculum Reform. *The International Journal of Learning*, 17(9), 403-424.
- Zhong, Guorong (Choong, Kok Weng). (2010). Han yu zuo wei di er yu yan de yue du li jie ce shi de xin jia po mo shi (Assessing Reading Comprehension in Teaching Chinese

as a Second Language: A Singapore Model). *Shi Jie Han Yu Jiao Xue (Journal of Global Chinese Language Teaching)*, 2010(3), 415-422.

Zhou, Jian & Huo, Bingkun. (2012). Jiao xue nei rong zhi shi de ding yi he nei han (The definition and essence of pedagogical content knowledge). *Xiang gang jiao shi zhong xin xue bao (Hong Kong Teachers' Centre Journal)*, 2012(11), 145-163.

Zhou, Qinghai. (Chew, Cheng Hai). (2009). *Bian dong zhong de yu yan (Language in change)*. Xin jia po: Lingzi chuan mei.

Zhu, Xinhua. (2005). *Neng li fa zhan dao xiang de yu wen ping gu yu jiao xue zong lun (A final collation on the "Ability-driven" language assessment, teaching and learning)*. Xin jia po: Zhong wai fan yi shu ye she.

Zhu, Xinhua. (2006). Jian gou "cu jin xue xi" de hua wen ping gu de ce lue (Strategies used in constructing assessment for enhance learning in Chinese Language). *Hua wen xue kan (Journal of Chinese Language Education)*.

Appendix A: Glossary of Assessment Concepts (Purpose of Assessment)

Formative Assessment is carried out during the instructional process to provide feedback to adjust ongoing teaching and learning in order to improve the students' achievement of intended instructional outcomes. It may involve informal methods such as observation and oral questioning, or the formative use of more formal measures such as traditional quizzes, portfolios, or performance assessments. Formative assessment entails observations which allow one to determine the degree to which students know or are able to do a given learning task, and which identify the part of the task that the student does not know or is unable to do so.

Assessment for Learning (AfL) is assessment that supports teaching and learning with the specific use of learner-centred approaches and strategies. For instance, teachers may identify gaps in student learning, and provide quality feedback for students to improve their work. AfL is used to redirect learning in ways that help learners master learning goals, and is primarily used for ensuring that the intended learning outcomes are achieved by students. For these reasons, it is a more 'extended' form of formative assessment, and is central to classroom instruction.

Summative Assessment is carried out at the end of an instructional unit or course of study for the purpose of giving information on students' mastery of content, knowledge and skills, assigning grades or certifying student proficiency. It is designed primarily to serve the purpose of accountability, or of ranking, or of certifying competence or achievement at a particular point. However, it is possible to use summative assessment in a formative way. For example, teachers may analyse students' performance in the class test and provide clear, detailed and descriptive feedback to help students bridge the gap.

Assessment of Learning (AoL) ascertains what students have learnt and functions as a means to judge if curricular outcomes have been met. It is used primarily for accountability purposes - grading, ranking and certification - to record and report what has been learned. For these reasons, it tends to be summative in nature and is usually carried out at the end of a unit, semester or year.

(MOE-OPAL, 2018)

Appendix B: Literature Reviews on Assessment Tension

Empirical Studies

S/N	Year	Study	Participants	Level	Country	Data Source	Data Analysis Method	Dimension	Tensions
1	1990	Taylor & Wallace	9 teachers	Sec	UK	Int, DA	CA (case study)	Two	Policy & practices
2	1991	Field	17 teachers	Pri	Canada	Int, Art	CA	Multi	Belief & practice
3	1995	Paechter	not reported	Sec	UK	Int, MO	CA (case study)	Two	Different purposes
4	1998	Monson	5 schools	High school	USA	Int, FGD	CA	Multi	Different purposes
5	1999	Bishop et al.	800 schools	Sec	UK	LS, Int, FGD	SA, CA	Multi	Different purposes
6	1999	Dunne	19 teachers	Sec	UK	Int	CA	Two	Policy & practices
7	1999	Ecclestone & Swan	11 lecturers	Higher Ed	UK	Int, LS	CA (action research)	Two	Policy & practices
8	2000	Grant et al.	16 teachers	High school	USA	Int, FGD	CA	Multi	Policy & practices
9	2000	McMillan & Nash	700 teachers	Pri+Sec	USA	LS, Int	CA	Multi	Belief & practice
10	2002	Lumley	4 teachers	General	Australia	Raters' think aloud data	SA, CA	Two	Theory & practices
11	2002	Reis & Villaume	199 preservice teachers	Teacher Ed	USA	LS, Int	SA, CA (case study)	Multi	Different purposes
12	2002	Sakonidis, Tsatsaroni & Lamnias	14 teachers	Pri	Greek	Int	CA	Two	Policy & practices
13	2002	Whittaker & Young	4 teachers	High school	USA	Reflections	CA	Two	Policy & practices
14	2003	Hornby	14 teachers, 441 students	Higher Ed	UK	LS	SA	Multi	Different purposes
15	2003	Lindsay & Lewis	982 teachers	Pri	UK	LS, Int	SA, CA	Multi	Different purposes

16	2004	Arkoudis & O'Loughlin	1 teacher	Pri+Sec	Australia	Teacher's story	CA (case study)	Two	Policy & practices
17	2004	Davison	12 teachers	Sec	Australia, HK	Int, LS, SR	CA	Multi	Cultural difference
18	2004	Hall et al.	4 teachers	Pri	UK	LO, Int, discussion	CA	Two	Policy & practices
19	2004	Hammerness	1 teacher	High school	USA	Int, LO	CA (case study)	Two	Policy & practices
20	2005	Asselin et al.	7 teachers	Pri+Sec	Canada	Int	CA	Multi	Policy & practices
21	2005	Cella	1 teacher	High school	USA	Int, auto	Critical discourse analysis	Two	Policy & practices
22	2005	Horwitz	NA	Pri+Sec	USA	Int	CA	Multi	Policy & practices
23	2005	Qi	388 teachers, 986 students	Sec	China	Int, LS	SA, CA	Multi	Different purposes
24	2005	Vandeyar	3 teachers	Pri	South Africa	Int, LO, DA	CA (case study)	Multi	Policy & practices
25	2005	Volkman et al.	3 teachers	Higher Ed	USA	Int, DA	CA (case study)	Two	Policy & practices
26	2006	Crook, Gross & Dymott	16 teachers, 16 students	Higher Ed	UK	Int, FGD	CA	Two	Policy & practices
27	2006	James & Pedder	558 teachers	Inf+Pri+Sec	UK	LS	SA	Multi	Belief & practice
28	2007	Scholtz	NA	Higher Ed	South Africa	Module test	SA	Two	Different purposes
29	2007	Strauss & U	14 lecturers	Higher Ed	New Zealand	Int	CA	Multi	Theory & practices
30	2008	King, Patterson & Stolle	3 teachers	Teacher Ed	USA	Int, LO	CA	Two	Policy & practices
31	2008	Maloley	2 teachers	Pri	USA	Int, LO	CA	Multi	Policy & practices
32	2008	Walls	8 teachers	Pri+Sec	Australia	Int	CA	Multi	Policy & practices
33	2009	Bonner & Chen	222 preservice teachers	Teacher Ed	USA	LS	SA	Two	Different purposes
34	2009	Gioka	9 teachers	High school	England	Int, LO, Art	CA	Two	Different purposes

35	2009	Webb & Jones	6 teachers	Pri	USA	LO, Int, discussion	CA	Multi	Belief & practice
36	2010	Arden	2 teachers	Sec	Canada	Int, LO, discussion records	CA (case study)	Two	Policy & practices
37	2010	Ashworth, Bloxham & Pearce	6 students & some staff	Higher Ed	UK	Int, LO, discussion records	CA (case study)	Two	Policy & practices
38	2010	Collins et al.	10 academics	Higher Ed	Australia	LS, Int, FGD	CA	Multi	Different purposes
39	2010	Deng & Carless	4 teachers	Sec	China	Int, LO, FGD	CA	Two	Belief & practice
40	2010	Eren	304 preservice teachers	Pri+Sec	Turkey	LS	SA	Two	Belief & practice
41	2010	Luk-Fong & Brennan	24 teachers	Sec	HK	Int	CA	Multi	Cultural difference
42	2010	Norashikin et al.	10 teachers	Pri	Singapore	Int	CA	Multi	Policy & practices
43	2011	Harman & McDowell	11 lecturers	Higher Ed	UK	Int	CA	Multi	Different purposes
44	2011	Johnson	4 teachers	sec	USA	Int, auto	CA	Two	Policy & practices
45	2011	Kostogriz & Doecke	3 teachers	Pri	Australia	Int	CA (case study)	Two	Policy & practices
46	2011	Volante & Beckett	20 teachers	Pri+Sec	Canada	Int	CA	Two	Different purposes
47	2012	Dreher	group of teachers	Pri	Australia	Int, LO, Surveys	CA	Two	Policy & practices
48	2012	Leach	472 students	Higher Ed	New Zealand	LS, Rubric self seessment	SA	Two	Policy & practices
49	2012	Lyon	11 teachers	Sec	USA	Int, LS, DA	SA, CA (case study)	Two	Different purposes
50	2013	Hardy	20 teachers, teacher aides,	Pri+Sec	Australia	Int	CA (case study)	Two	Policy & practices
51	2013	Kim et al.	91 teachers	Sec	Singapore	LS, FGD, RW	SA, CA	Two	Policy & practices
52	2013	Petcovic et al.	17 teachers	Higher Ed	USA	Int	CA	Two	Belief & practice

53	2013	Yin	1 teacher	Sec	USA	Int, LO, DA	CA	Two	Different purposes
54	2014	Suurtamm & Koch	42 teachers	Pri+Sec	Canada	Int, DA	CA	Multi	Different purposes
55	2014	Tan & Leong	2 teachers	Sec	Singapore	Int, DA	CA	Two	Policy & practices
56	2014	Thanh Pham & Renshaw	2 lecturers, 250 students	Higher Ed	Vietnam	LS, Int	CA, SA	Multi	Cultural difference
57	2015	Fleer	11 teachers	Pri	Australia	Multiples	CA	Multi	Different purposes
58	2015	Hopfenbeck, Flórez Petour & Tolo	98 educators	Pri+Sec	Norway	Int	CA	Two	Policy & practices
59	2016	Pratt	12 teachers	Pri	UK	Int	CA	Two	Policy & practices
60	2016	Pinto	Groups of educators	Pri+Sec	Canada	Various	CA, SA	Multi	Policy & practices
61	2017	Sadeghi & Rahmati	16 students	Pri+Sec	Iran	Cambridge Assessment,	CA, SA	Multi	Theory & practices
62	2017	Cooper, Susan Elizabeth	NA	High school	USA	Int, think-aloud protocol	CA	Two	Purposes

Reviews or conceptual papers

S/N	Year	Study	Level	Country	Data Sources	Dimension	Tensions	Theoritical Framework
1	1989	Troman	General	UK	Multiple policy papers	Two	Policy & practices	Yes
2	1994	Delandshere & Petrosky	Teacher Ed	USA	National Board Certification	Two	Theory & practice	Yes
3	1994	Taylor	General	USA	NIL	Two	Different purposes	Yes
4	1995	Black	Secondary	UK	Education Reform Act	Two	Policy & practices	Yes
5	1996	Delandshere	Teacher Ed	USA	NIL	Multi	Policy & practices	Yes
6	1997	Hill et al.	Secondary	Australia	Victorian Certificate of Education	Two	Policy & practices	No
7	1998	Firestone	Secondary	England & Vermont	Testing policies	Multi	Policy & practices	Yes

8	1998	Tierney et al.	Secondary	USA	Portfolio assessment	Two	Theory & practice	Yes
9	2000	Green & Brennan	Higher Ed	Australia	Vocational education and training	Multi	Policy & practices	No
10	2002	Morgan et al.	Secondary	UK	GCSE Maths	Two	Policy & practices	Yes
11	2002	Weller	Higher Ed	UK	UK Open University web-based course	Two	Policy & practices	No
12	2004	Palmer	Higher Ed	Australia	Deakin University School of Engineering	Two	Different purposes	No
13	2004	Timma	General	Australia	Vocational education and training	Two	Policy & practices	Yes
14	2005	Harlen	General	UK	Multiple reviews of research	Two	Different purposes	Yes
15	2007	Cooper & Heinze	Higher Ed	UK	Bachelor of Science programme	Two	Theory & practice	Yes
16	2007	McKay & Brindley	General	Australia	Multiple reviews of reforms and research	Two	Policy & practices	Yes
17	2011	Langfeldt & Kyvik	Higher Ed	NIL	NIL	Multi	Different purposes	No
18	2011	Myran & Clayton	General	USA	US Doe School Leadership Program	Two	Different purposes	Yes
19	2012	Yu & Frempong	General	USA	NIL	Two	Different purposes	Yes
20	2014	Exley & Chan	General	Australia	Australian Curriculum English Version 5.0	Two	Policy & practices	Yes
21	2015	Birenbaum et al.	General	Multiple countries	NIL	Multi	Different purposes	Yes

Appendix C: Sample invitation email to participants

XX:

您好！之前得校长的同意，让我到学校收集一些我个人学位研究的资料，我非常感谢大家给我这个机会。给你添麻烦了，真不好意思。

我想先请你看看一些资料（请看附件），然后才正式进行访谈。这包括我的研究背景资料和访谈问题。假如你没有什么时间，就先看访谈问题，因为研究背景资料可以等到我们见面时才告知。

假如可以的话，请先安排一、两个你可以配合的时段进行访谈，我会配合你给的时间。让你费神，还望见谅。谢谢你。

祝：工作愉快！

广通 敬上

Appendix D: Attached email documents

Basic Research Information for Participants

(A) Basic information:

S/N	Information	Details
1	Researcher	Chan Kwong Tung (陈广通)
2	Contact details	Email: Kwongtung2011@gmail.com Mobile: 98196353 Office: 68796563 Home: 67492249
3	Programme	Doctor in Education (EdD) Programme (enrolled in 2012) from Curriculum, Teaching and Learning Academic Group, NIE, NTU.
4	Title of research	A Phenomenographic Study of Singapore Chinese Language (CL) Teachers' Experience of Assessment Tension (AT) 以现象描述分析学看新加坡华文老师的评价张力经验
5	Supervisor	Assc Professor Tan Heng Kiat, Kelvin
6	No. of teachers involved	9 CL teachers
7	Data collection method	Personal interviews (audio recorded), each session may take up 30-90 minutes

(B) Research summary:

<i>Item</i>	<i>Focus</i>
Research Topic	Teachers' experience of assessment tension in practice
Research Problem	Teachers' experience of assessment tension is not well recognized as an important educational problem and thus is poorly understood
Purpose Statement	To study the qualitatively different ways in which teachers experience assessment tension
Research Question	What are teachers' qualitatively different ways of experiencing assessment tension?

(C) Abstract of proposed research:

Unlike many past studies, this proposed study intends to explore assessment issues from the perspective of the "experiencers" on the ground by employing phenomenography. This will help further unpack the actual meaning of "assessment" for all stakeholders of education to understand its meaning and its implication for use in

the classrooms. Although there are many cases of assessment tension (AT) reported by scholars in previous studies, they have yet to explore “the black box” of teachers’ experience of AT, especially the voices from the “experiencers”. This is the core educational problem in my context which past studies and scholarships is not able to address satisfactorily at the moment.

The overarching research question is: What are the qualitatively (or distinctively) different ways in which teachers experience AT?

Teachers experience AT at different levels when teaching and assessing students’ learning when there is a call for assessment change in practice. In this research, **AT is defined as “teachers’ feelings of struggle with competing considerations in their assessment practices”**. I caution that a continuous failure to recognise and address AT could continue plaguing a wider application of teacher empowerment and professionalization of teaching in practice. When teachers experience AT, it dramatically affects teacher empowerment. When this happens, teachers’ assessment decisions depend on other external factors rather than on their professional judgement. Teachers will face the problem of inconsistency, uncertainty and even compromise when making assessment decisions. As a result, a series of negative issues associated with teachers’ work related outcomes, such as low job satisfaction, poor work and organisational commitment, which will have a direct impact on personal motivation, classroom improvement and school effectiveness.

Looking into teachers’ experience of AT provides a dynamic perspective of understanding and conceptualizing assessment which diverge from the more distinctive and quiescent perspectives of understanding assessment. By investigating Chinese Language (CL) teachers’ of AT, this subject-specific, single context study will add richness to the general AT studies which usually encompass various subject disciplines. By employing phenomenography, the interconnectedness and comprehensiveness of the AT experienced by teachers could then be further explored. This will subsequently help us find ways to support teachers to carry out sound assessment with minimum tension. Phenomenography is able to capture and explore teachers’ collective experience of assessment tension (AT) in terms of their distinctive differences (or variations) and their interrelatedness by employing the “maximal variation sampling” strategy that allows this study to purposefully select teachers who are different in their characteristics and backgrounds. Data will be collected from the personal in-depth interviews.

The interview data analysis is to make sense of the particular expressions in terms of the individual, the group of participants, as well as their interrelated relationships. Eventually, after the transcription stage, it is to develop a set of categories of description (the qualitatively different ways of experiencing AT) that characterise the distinctive differences in how the AT has been experienced by the teachers. Finally, to establish a full progression (or hierarchy) of the different ways of experiencing AT, an “iterative process” may be used to produce the final version of descriptions of categories.

Key words: Assessment tension, Chinese Language, Singapore Chinese teachers, phenomenography

(D) Interview procedures:

- All interview sessions will be conducted in a pre-arranged room which will be conducive for discussion.
- All interview sessions will be recorded using two digital voice recording devices and transcribed into first Chinese (since the participants are all CL teachers) and then translated into English.
- All interviews will be arranged according to the working schedule and teaching timetable of the teachers involved so that there will be minimum disruptions to their daily routines.

(E) Anticipated benefits and risks to human subjects:

Teachers and schools participating in this proposed study may benefit in the following ways:

- The identified teachers' assessment practice and their considerations when making assessment decisions may help inform future professional development and school policy making in assessment.
- Teachers and school leaders may start to widen the foundation of common comprehension and meaning of the assessment tension and find effective ways to address and resolve this practice problem.
- Participants and school leaders will be invited to attend a sharing session by the researcher after the degree has been awarded. However, no direct personal benefit is promised in this participation.

Although there should be minimum risk arising from the participants and schools, the proposed study will take all necessary precautions to reduce the predictable research procedures that might cause harm to the participants by ensuring the followings:

- Ensure that all data collected regarding the teacher's assessment experience interviews are not used for any judgmental purposes and will not be reported to the school leaders and respective reporting officer for performance grading or ranking.
- Announce that the confidentiality of the data and information of the participants in the thesis will be ensured.
- Make clear to the participants on the intention for collecting data and to ensure them that the study will have the benefit of the participants in mind.
- Since all interviews are audio-tapped, the participating teachers can request to delete any section of the sound recording if they do not feel comfortable with the recording.
- Participation is voluntary. Although all participating teachers have given a written consent to participate in the proposed study, they can withdraw at any time from the participation without giving any reasons.

(F) Ethics review and standards:

For more information on ethics review and standard, please visit NTU IRB at <http://research.ntu.edu.sg/GuidelinesnForms/Pages/default.aspx> or contact Ms Germaine Foo at 65922495; email: IRB@ntu.edu.sg.

Participant/Teacher Consent Form

Please read this consent agreement carefully.

You will be invited to participate in the social and behavioral sciences research carried out by the researchers at National Institute of Education, Nanyang Technological University. As your participation is voluntary, we seek your consent to participate in these studies.

What would you be asked to do? In these studies, you will be asked to report or share your experience, conceptions, beliefs, attitudes, opinions, behavioral preferences, and habits.

Is my participation compulsory? Your participation is voluntary. In all studies, before the study begins, you will be fully informed of what you will be asked to do and be allowed to ask any questions at any time that you might have about the study. You may refuse to participate, discontinue participation, or skip any part of the task you do not wish to work on at any time without penalty. You are also free to withdraw from these studies at any time. Your decisions will not affect your assessment by your supervisor, performance or status in your school.

What would you gain? Your participation in the research will eventually provide you with the opportunity to acquire first-hand knowledge of the research topic, process and findings. The relationships among the different stakeholders of education in this research context could also be gained.

What are the risks of this research? The Internal Review Board (IRB) of Nanyang Technological University will review the studies and this form will be relevant only when the Board has approved the studies and determined that there are no anticipated risks, beyond those encountered in daily life, associated with participating in these studies.

How will my privacy be protected? Your responses will be kept private. Only the researchers of the pertinent project will have access to your responses. In the event of publication of this research, no personally identifying information will be disclosed.

Who should I contact for questions and clarifications? You are encouraged to contact **Chan Kwong Tung** for questions and clarifications. You may also contact the Institutional Review Board via email at irb@ntu.edu.sg, or through phone call at 65-65922495.

If you consent to participate in these studies, please sign and return this form to NTU.

I have read the above information and give consent for my participation in the research titled "A Phenomenographic Study of Singapore Chinese Language (CL) Teachers' Experience of Assessment Tension (AT)" conducted by Chan Kwong Tung

Name of participant/teacher: _____

NRIC/Passport Number: _____

Signature: _____

Date: _____

Appendix E: Interview Protocol

Title of research: A Phenomenographic Study of Singapore Chinese Language (CL) Teachers' Experience of Assessment Tension (AT)

以现象描述分析学看新加坡华文老师的评价张力经验

Time 时间	
Date 日期	
Venue 地点	
Interviewer 受访者	
Interviewee 访问人	

The purpose of this research is to find out the qualitatively (or distinctively) different ways in which Chinese teachers experience assessment tension (AT). The main data will be collected via the semi-structured interviews with teachers

(A) Three sets of Overall Interview Questions:

(1) What is your assessment practice and how do you carry out assessment?

（在你的教学中使用过什么评价的工作？你如何进行这些评价工作？）

- When do you use assessment and for what purposes? （你在什么时候进行评价工作？这些评价工作有何作用？）
- How often do you carry out assessment? （你多久进行评价工作？）
- What kind of assessment do you use in your classrooms and why? （在教学中，你都用哪一类评价工作？你为何使用这些评价？）
- Where were your assessment tasks from? Which types has more and why? （你的评价工具都来自何处？哪一类评价比较多？为什么？）

(2) In your assessment practice, what are your considerations: such as time, resources, accountability, students' needs etc?

（在你进行评价工作时，你有何考量？例如：时间，资源，问责问题，学生的需要等）

- Why do you use these assessment and not others? （你为何用这些评价？为何不用其他的？）
- How do you feel when carrying out these assessment? （在你使用/进行这些评价时，你有何感觉？）
- If you were given the freedom to design assessment, how would it look like? （假如你有绝对的自由，没有任何的限制，你会设计怎样的评价？和你现行的会有分别吗？会有何分别？为什么？）

(3) How do you deal with these different considerations?

(你会如何处理或权衡这些在评价工作上的不同考量?)

- Do you think there are better assessment practices?(你认为有更好的评价方式吗?)
- Do you seek other alternatives? (你有尝试过其他的评价吗?)
- How do you react to these considerations? (对于这些不同的考量, 你有何反应?)

(B) Sub-questions for clarification and elaboration:

- Could you explain that further? (你可以再解释清楚吗?)
- Could you give me some examples? (你能提供一些例子吗?)
- What do you mean by that? (你这样讲是什么意思?)
- By saying.....do you actually mean/imply that.....? (你这样讲, 是不是意味着.....)

Appendix F: Sample of preliminary categories of description

AT	Sets of AT	Category 1	Category 2	Category 3
1	Assessment for learning vs preparing students for national examination	Lack of time to prepare	Lack of resources, materials to get students ready	Student-centred assessment
2	Current assessment practice vs their ideal assessment	Technical issues of assessment	Language assessment approach	Subvert traditional language assessment
3	Assessment literacy vs competencies in carrying out quality assessment	Lack of time and opportunity to improve	Lack of assessment knowledges, resources, strategies and tools	Overturnd the conventional paradigm of assessment literacy
4	Formative assessment vs summative assessment	The amount of assessment needed to be carried out	The negative effect of assessment	The unbalance assessment practice



S/N	Dimension of variations	Category 1	Category 2	Category 3
1	No. of factor(s) surfaced	Single	Multiple	Multiple and complicated
2	Affective level	Receiving	Responding	Valuing / Organizing

	AT experienced (The “what”)		Three different ways of experiencing AT (The “how”)		
AT	Understood AT in terms of	Distinctive AT experienced	Fatalistic (宿命论的)	Responsive (响应的)	Awakening (觉醒的)
1	Stakes	Assessment for learning vs preparing students for national examination	Lack of time to prepare students for the national high stake exam	Lack of resources, materials to get students ready for national exam	Striving toward student-centred assessment
2	Theory and practice gap	Current assessment practice vs teachers’ ideal assessment	Technical constraints of assessment in practice	Differences in language assessment approach	Subvert traditional language assessment practice
3	Assessment literacy	Assessment literacy requirement vs teachers’ competencies in carrying out quality assessment	Lack of time and opportunity to improve assessment literacy	Lack of assessment knowledges, resources, strategies and tools	Overtured the conventional paradigm of assessment literacy
4	Purposes	Formative assessment vs summative assessment	The overwhelming amount of assessment needed to be carried out	Cautioned the negative effect of assessment	Re-think about assessment effect in the long run
Dealing with AT			Passive executer (消极的 执行者)	Inquisitive practitioner (多问 的实践者)	Rational re-conceptualizer (理 性的重新概念化者)

Summary of outcome space (three different ways of experiencing AT)			
Conceptions of AT	Fatalistic	Responsive	Awakening
(a) Focuses on	Very obvious factors which were not within teachers' sphere of control. These included lack of time to prepare quality assessment, technical issues of assessment, lack of opportunity to improve teachers' assessment literacy and the large amount of assessment tasks in schools.	Common and explicit factors which somewhat within teachers' sphere of control. These included assessment resources and materials, language assessment approach, assessment knowledge and strategies, the negative effect of assessment.	Not common and obvious factors which were not so substantive but philosophical in nature. Teachers re-think the functions of assessment, subvert traditional assessment approach, overturned the conventional paradigm of assessment literacy and re-conceptualized the effects of assessment in the long run.
(b) Affective level	At the receiving level, teachers were aware of and could acknowledge the phenomenon. They were sensitive to these AT but stayed rather passive in response to their experiences.	At the responding level, teachers were more interested in the AT and would react to this phenomenon in an active manner. They may complain, question and even probe further on the current assessment practice and approach.	At the valuing / organizing level, teachers were deciding the worth and relevance of their assessment practice and experience. They re-examined the meanings behind these values and were willing to re-conceptualize their belief and practice toward assessment.

Appendix G: Proposed conceptions of assessment tension

Version	Conception A	Conception B	Conception C
1	Passive	Respondent	Awakening
2	Passive	Respondent	Reflective
3	Disempowering	Engaging	Empowering
4	Disempowering	Responsive	Empowering
5	Fatalistic	Responsive	Empowering
6	Fatalistic	Responsive	Propulsive
7	Static	Responsive	Propulsive
8	Static	Potential	Propulsive
9	Stasis, inert	Potential	Propulsive
10	Inertial	Potential	Propulsive
11	Inertial	Accelerating	Propulsive
12	Inertial	Exertive	Propulsive
13	Passive	Exertive	Propulsive
14	Passive	Responsive	Propulsive

Appendix H: Original Extracts in Chinese

Extract	Teacher	Original transcription
1	C	<p data-bbox="523 309 1318 730"><i>Teacher C: I never thought of formative or summative assessment. I don't really understand these terms. Actually I think my practice is more 'exam-oriented'. Because I need to be accountable to students' parents. That is when it comes to exams, school, students will need to know these exam scope and requirement and so we need to teach them beforehand, that is to prepare them well for exams. Therefore, I give them these. Yes, yes that's why I every time, that is these assessment, generally are designed according to the exam, directly related. Yes, I will not, I I seldom do anything beyond exams, those task-based ones.</i></p> <p data-bbox="523 775 1442 1137">我没有去想过形成还是终结。这个我搞不清楚这个。其实我觉得我，我的要求是比较 exam oriented 的。因为我要向学生的家长交代的。就是到了他们考试的时候，学校额，学生要考试的这些范围跟他们要能够做的这些要求啊，我都要在他们考试之前呢就是教他们，就是让他们备考啦，就是准备好了去考试。所以我给他们的这些。对啊，对 that's why 我我每次啊，就是这些评价一般上呢我都会设计到跟他们的考试啊其实有，有直接关系的。对，我不会，我我也很少去做很多就是这些跟考试以外的一些额考试以外，任务型</p>
2	I	<p data-bbox="523 1153 1318 1574"><i>Er, assessment for learning, I think teachers should accept it with a positive thinking. In other words, we should, firstly, not waste time and secondly, we should not be asked to see the effect immediately, I want to see the teaching effect. How do we not face with the stress of exams? We must overcome this challenges, use this assessment for learning to let your teaching to be more fantastic, to be more effective and at the end their national exams will not be poor. This is my conclusion; they will not score poorly. It allows me to prepare my students for exams more confidently. I think that's really a way to helping me, my students to do well in the exams.</i></p> <p data-bbox="523 1619 1430 1856">嗯...教学评价我觉得教师应该用正面的这个思想来接受啦。换句话说我们，第一我不要浪费时间，第二我不要立竿见影，我要看到教学效果，怎样都不能够面对呢那个考试的压力对不对？你必须克服，用这个教学评价让你的教学更加精彩、更加有效，最后你不会考差啊！这个是我的总结，不会差，真的不会差。它使我更加有把握地考出好成绩。</p>
3	I	<p data-bbox="523 1870 1318 2016"><i>My challenge in assessment is how to get at least 50% distinction in exams. Yes, to do well in the exams. At the end of the day, the student can use Chinese as a career or profession, or related to language, you know or be a DJ in a</i></p>

		<p><i>radio station. That is big, that is my biggest challenge. I have not achieved that yet. For example, Desmond Koh. For example, Huang Shuangxi and so forth. The student can use Chinese as a tool for their work in future. They have to be among the best in Chinese. Then I think it will be a big success in using assessment.</i></p>
		<p>哦最大的挑战当然就是说最少最少要 50percent distinction 啦。考好成绩。最后他以华文作为教学事业或者作为工作语言或者是什么或者成为电台里面的。那个很大，那个是我最大的挑战，我没有做到。譬如说许振荣啊、譬如说黄双喜啊什么，他最后能够以华文来作为谋生工具，我想那是华文里面的精英翘楚。我觉得那里也是一个很大的成功。</p>
4	A	<p><i>Language assessment, actually I felt that for each student, there are great differences among their language ability. Some are good at speaking. So if for me, I felt it is, our national assessment in language is looking at listening, speaking, reading and writing holistically. But I felt some, he could be one who is just good at speaking. So is there a possibility that his language ability is mainly exhibited in speaking? Is like there are many who studied Chinese but at the end not many turned out to be a writer.</i></p>
		<p>语文的评价哈，其实我觉得对每个学生来讲，他们的能力是有差异的，有的人擅长讲。所以如果是我的话，我觉得，就是说，我们的，我们的评价好像比较要的是一个全才。就是听说读写的。但是我觉得有一些人，他可能就是一个，他就是口才好。那有没有一个可能性就是说，他的语文能力他就是主要更多就体现在他的口语方面？好像我们中国人很多都学华语，到最后成为作家的没有几个。</p>
5	E	<p><i>For example, in Paper 1, I think there is no need to test a functional writing and a composition. The two parts are testing the same skills. Because one writing skill, by only test one component, we actually would be able to know where is his skill.</i></p>
		<p>比如说试卷一，我觉得没有必要又考他一个实用文又考他一个作文。因为一个写作技能哈，只要考一个项目，其实是已经能够知道他的能力在哪里</p>
6	D	<p><i>The current assessment approach is on the listening, speaking, reading and writing. There can be variations in between but eventually still on these skills. If you really give me absolute freedom to think about my ideal assessment, I think language assessment should include argument as a mode. Because to argue, that is to let them see the difference in viewpoints and there will be ‘thought collision’, those are</i></p>

		<p><i>inner thinking which need preparations. For example, I can give a current hot topic and divide the students into groups for them to prepare the information needed. They need to do research and present them into sentences and do presentation. They cannot any how say and thus the sentences must be refined. So during this process, we can assess many abilities. How do you cooperate? How do you do research? How to search on the Internet. And how to write it out?</i></p>
		<p>评价方式可能是可以在这个基础上，听说读写这个基础上，可以是花样百出，但是你最终你其实的你最终还是落到那几点上面啦。你比如说如果说你真让给我绝对的这个自由哈，我觉得语文当中辩论是挺好的一个方式。恩，因为辩，因为我现在挺喜欢给我学生看这个有关辩论的一些话题，就是双方不同的观点的时候，他会有思维的碰撞，就是他内在的思维，你你会听对吗？如果说有这个可能的话，如果我们学生的水准又在那儿的时候，其实也不存在在不在那儿，其实我觉得辩论是可以准备的。比如说我们给他一个话题，对吧？给他一个当下热门的一个话题，然后分两组咯，对不对？然后每一组他要去准备资料。他要去搜集这些资料，他要把他的资料整理成一段话，他的话就必定不是，不可以乱说的，一定要经过修饰的，他才可以说出来，所以这个，前面的这些过程你其实考察了很多能力对不对？他怎样合作啦，他怎样搜集资料啦，他怎样去网上找跟这个主题相关的东西啦，他怎样，找完这些东西他要把它变成写的。</p>
7	B	<p><i>So for task-based assessment, I think it is very close to students' lives and they really are using the language in an authentic setting. So students need to learn to use. And use it in their actual lives. I think the listening, speaking, reading and writing skills need to be tested. How do you test? They must be able to use, you must be able to use it in an integrated manner and give them an authentic environment to activate their language ability to complete the task. That is the best evidence.</i></p> <p>所以我这个是一个 Task-based assessment，我觉得它非常贴近学生的生活，那么学生是的确是在正式语境里头碰到这样子的情况。所以学生他不单只是学了，他还是必须要用。用也是跟他的生活有很大的联系跟关联。而且我觉得听说读写的这个能力呀，它必须给，你怎么考查他，他必须要能够用。你，他必须要综合运用，然后给他一个实际的环境，让他去去调动他听说读写的能力，来完成这个任务，这个是最好的证明。</p>
8	E	<p><i>I have this thinking because I found that when students learn, whether it is language learning or not, actually there is this 'inner quality' which, for example, this 'keep on fighting' spirit. Because of this quality, so this is also an influencing factor for him to learn the language. So when you assess or</i></p>

		<p><i>help him to improve this inner quality so that he can transform it into language learning, I think this is it. Students must have certain inner quality that allow them to fully capitalise on the assessment given by teachers. If not, this assessment would be useless.</i></p>
		<p>因为我发现哈，学生的学习是，不管是语言还是什么，其实他跟他的嗯一种内在的一种品质是有关的。所以可以譬如说他能不能够屡败屡战。所以这个它也是一些影响他的语文学习的一些条件。你评估或者是帮助他提升这些内在的一些品质的话，然后能够帮助他把这些内在的一种品质转化到学习上的话，我觉得是。学生他的内在素质并没有使他去配合这个评价工具的话，这个评价工具就没用。</p>
9	G	<p><i>Language teaching and assessment has to be flexible and multi-faceted. So I would hope this mini novel teaching can bring what kind of reading experience to the students, or let them associate with something in their lives. You mobilize their emotions and then you allow them to write a paragraph. They would be able to write. For this test, it is beyond just language assessment. And it is really practically assessing how the child's can allow his emotions to write into words, about how to express himself. This is very touching.</i></p> <p>它不一定是很硬性的，但是我比较会希望说这个微型小说带给学生什么样的读故事的经验，或是会让他联想到生命当中有什么东西。你调动他们的情感，然后你让他们写出一段来，他们就真的可以，这个时候你测试的这个超越了语文界的。而是真的很实在地测试小孩怎么样把情感写成文字，怎么样表达，那么会很感动的。</p>
10	B	<p><i>I have these Sec 5 normal students. Oh, that is at the beginning, I told them the test marks are not important, what is important is the corrections. Then they didn't listen. The second time I told them again but they refused to listen again. And so the third time. Therefore, I felt that I was wasting their time, I was doing things which were thought of not important. Then, I didn't do it for every question. I still return their tests and of course in order to make my teaching more effective and here if I cannot change the students, and they don't take your advice and think that it is not important, of course I have to make changes. So I will perhaps reduce descriptive feedbacks. So from those important ones, for example those ABCD questions, I will just tell them the correct answers. But I used to explain and explain. So if they don't listen once or twice, then I will just provide the correct answers.</i></p>

		<p>我对那种中五 normal 的学生呐，啊就是开始第一次跟他说分数不重要，重要对视订正，然后不听，第二次也不听，第三次也不听变成，我就感觉到浪费他的时间，在做他们觉得不重要的事情。后来我就啊，我也啊不每的题目，就是我还，当然我就是让我的教学更有效的话，然后这边改变不了学生的话，学生也一直不买你的账，然后也觉得不重要，当然我就做一些调整喏。所以，所以这边可能在和他们对答案的时间当然会相对减少啦，所以，啊从可能一些重要，譬如说一些 ABCD 的那个的 ABCD，你选错只要有答案就可以了嘛，很简单。可是以前我会说跟你解释解释，那解释学生不听的话，然后一两次不听，可能后来也跟他们这样做的话，就是给你一个答案喏。</p>
11	C	<p><i>And because of this assessment task, I had an argument with a parent. Because this parent reprimanded me for giving her child the "I love kitchen" project. But that student did not do. Then I told her the child did not do the project but I have to key in the CA marks. So no marks lah. At the end, the parent was not happy. That was rather complicated because there were other things and argument beforehand. Yes, then I told the parent there will be no mark and so on and I forgot the details because that was a while ago. At the end, she came to reprimand me during the parents-teachers' interaction day. Then, this parent is really, asked me why I gave her child this project task, she was very rude. She said "Why do you give this kind of assignment?" Then she placed her mobile phone on the desk, "what are you doing? The task you gave was so very tough. How do you expect my child to do?" Ha, I was thinking, this is considered tough? Many students had done that in the past. There is not much requirement in the project, just so simple with some steps. I didn't expect you to cook or stir-fry or fry food. You can do a sandwich-making. She said I had given a tough job. What do you expect me to do? How to do? At the end, I didn't want to argue with her, so I got Michael (the level head) to talk to her</i></p> <p>也因为这个作业，上次我跟这个家长就是发生过这争执。因为那个家长他就指责我就是说这个，我让他孩子做的是那个“我爱厨房”的那个系列的那个作业。然后嗯他孩子没有做吧，然后我有告诉他他没有做，然后没有做的话我 CA 我要算分，就没有分啦。后来他，他就不高兴啦。那其实也挺复杂的，因为之前还有一些其他的额…其他的一些争吵。恩对。然后后来我就跟他讲说这个是没有分，之类的我忘记了，因为好一阵子前了，后来有一次他就在家长日的时候他来找我嘛，我知道他，他就是来责问我啦。然后他，这个家长真的是，问我为什么给他（孩子）这样子的作业，然后额他很没有礼貌叻，他额他说“你为什么给学生这样的作业？”，然后他就拿他的手机出来啊摆在桌上叻，“你做什么？你给的任务这么艰巨！你要我的孩子怎样做？！”哈？我就在想，这样叫艰巨啊？过去已经有很多学生做过了。也没有什么</p>

		<p>要求，就是这样简单，就是讲一个步骤，我又没有要你要煮啊要炒啊要煎啊，你做个三文治也可以。他说你给我孩子的任务这么艰巨啊，你要我怎样，怎么样怎么样做。后来我不要跟他吵啦，因为我叫我叫 Michael 说你跟他谈，我不要跟他谈，从此之后我不跟他谈。</p>
12	I	<p><i>So if, I was thinking, I was thinking, if our teaching and assessment can involve parents to give feedbacks, I think it will be rather good. But this idea will make many teachers 'jump'. If assessment can involve parents, that is come in to play a supervision role, I think it will be better. But teachers may not..... Sometimes, if we let teachers, for example, if they are bolder enough, that to say, ah, "Dear parents, I am the CL teacher of 1E1, if there is any gap in my teaching, do inform me directly". Ah, then if parents have any feedback, that is also part of assessment. I am not sure if teachers can accept this? Interactive. Ah, I did try. I would inform parents that for this term, I have given ten assignments and so on. Your child has only handed in three assignments. Could you help me with the rest? I did this, so if the parents were able to help me take note of their children's learning, then they will be more serious.</i></p> <p>所以如果我在想哈，我在想，如果能够教学评价呢家长能够给一些反馈意见我觉得是挺好，不过呢你这个意见推出来很多老师会跳起来。如果评价能够把家长，额就是说进来扮演就是一点点督促的角色啦我觉得会更好啦，但是老师可能 tsk...不是...都能够做得到。有时候叫你老师，我假设啦，就是说老师如果胆子大一点哪，就是说“诶，亲爱的家长，我是今年 One E One 的华文老师，我在教学工作里面呢如果有什么疏漏的话请你直接跟我说。”啊那么如果是家长有反馈的话这也是评价的一个部分。我不知道老师能够接受吗。额，我有试啦。我会告诉家长，就是说这个学段哈我给了十篇作业什么，贵子弟呢只交三篇，你可以不可以帮我拿？我有这样。所以他如果是说帮我注意的话这个孩子会比较认真啦。</p>