

English Teachers' Reception of the New Assessment System.

Terry Zammit

A dissertation presented to the Faculty of Education in Part Fulfilment
of the Requirements for the Degree of Master in Teaching and Learning
in English with Second and Foreign Language Teaching and Learning.

June 2020.



L-Università
ta' Malta

University of Malta Library – Electronic Thesis & Dissertations (ETD) Repository

The copyright of this thesis/dissertation belongs to the author. The author's rights in respect of this work are as defined by the Copyright Act (Chapter 415) of the Laws of Malta or as modified by any successive legislation.

Users may access this full-text thesis/dissertation and can make use of the information contained in accordance with the Copyright Act provided that the author must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the prior permission of the copyright holder.

Dedication

I would like to dedicate this dissertation to my grandparents and my mother, who have been of constant support over the last 5 years at University.

Acknowledgments

First and foremost, I would like to begin by thanking my dissertation supervisor, Professor Doreen Spiteri, for constantly being available to discuss the content of this dissertation, to provide feedback, guidance, and support when it was most needed. I would also like to thank her for being a constant figure in my development into an educator.

I would like to thank Ms Clarissa Padovani, an Education Officer for English, for acting as an intermediary between me and the participants of this study, and also for helping me to access information when it was needed.

I would also like to thank my dear friend, Andrew Spiteri, for the many hours spent discussing statistical formulae and for teaching me to appreciate the power of numbers.

Finally, I would also like to acknowledge the many great teachers I have had throughout my time as a student at the University of Malta, and elsewhere, who undoubtedly had an influence on my choice to enter this Master's course.

Abstract

The Maltese educational sector is currently undergoing large-scale reforms with changes being implemented in the curriculum and various syllabi, the pedagogies used to educate Malta's youth, and the ways used to assess students' learning. One of these reforms consists of how students' Secondary Education Certificate examinations in English (Language and Literature) is assessed. The reform posits that, as of 2020, students will begin to be assessed via School-Based Continuous Assessment in the forms of various tasks for the English Language examination and assignments for the English Literature examination. This study aims to examine how the teachers are responding to having to take on the role of assessing segments of these qualifications as parts of their professional lives. The Literature Review identifies various factors which are associated with how teachers may respond to such reforms, with these having then been used to create an online questionnaire. The areas under which the factors fall are divided into three broad areas: what the teachers think about the reform, how they think the reform will impact their working lives, and their efficaciousness when it comes to handling the reform. Each one of these also represent a research question in the study. The study intended to follow a mixed methodology approach, yet various factors led to the qualitative aspect being abandoned in favour of a more quantitative approach. The study concludes that, whilst there are a number of areas which are indicative of a positive reception on the teachers' ends, there are also a number of areas which are being perceived negatively. This mixed reception and its implications for stakeholders are discussed.

Table of contents.

DECLARATIONS BY POSTGRADUATE STUDENTS	i
Dedication	ii
Acknowledgments	iii
Abstract	iv
List of figures	viii
Chapter 1: Introductory chapter.....	1
1.1: The Maltese context.	1
1.2: Statement of the problem.	2
1.3: Frameworks considered.....	3
1.4: The Local Research Context.	6
1.5: Research rationale and research questions.	8
Chapter 2: Literature review.....	11
2.1: What is assessment reform?.....	11
2.1.1: <i>Defining assessment and reform.</i>	11
2.1.2: <i>Reasons for studying reform in the English SEC examination.</i>	12
2.2: Outline of the literature review.	12
2.3: A review of the reform.....	13
2.3.1: <i>The role of the public and culture in assessment reform.</i>	13
2.3.2: <i>The pressure for reform and the Maltese response to this.</i>	15
2.3.3: <i>The political implications of the proposed reform.</i>	18
2.3.4: <i>The Influence of Formative Assessment.</i>	20
2.3.5: <i>Factors affecting the successful implementation of the reform.</i>	22
2.3.5.1: <i>Teachers’ Conceptualization of Assessment</i>	22
2.3.5.2: <i>Issues of Reliability and Validity.</i>	23
2.3.5.3: <i>The washback effect.</i>	25
2.4: Factors and variables determining teachers’ receptivity.	27
2.4.1: <i>A definition of ‘receptivity’.</i>	27
2.4.2: Teachers’ knowledge and beliefs.	27
2.4.3: <i>Cultural expectations and their impact on assessment reform receptivity.</i>	31
2.5: The professional development issue.....	35

2.6: Self-efficacy and assessment literacy.....	39
2.6.1: <i>Definitions of self-efficacy and assessment literacy.</i>	39
2.6.2: <i>What does it mean to be “assessment literate”?</i>	40
2.6.3: <i>Integrating assessment literacy and self-efficacy.</i>	42
2.7: Conclusion.	43
Chapter 3: Research methodology	44
3.1: Research methodology and research design.	44
3.2: Selection of sites and participants.....	45
3.3: Instrumentation.	46
3.4: Data collection and recording.....	47
3.5: Methodological assumptions.	48
3.6: Limitations.....	49
3.6.1: <i>Limitations of removing the qualitative aspect from the dissertation.</i>	49
3.6.2: <i>Limitations of the online survey as a data collection tool.</i>	49
3.7: Establishing credibility.	50
3.8: Ethical considerations.	51
3.9: Restatement of the purpose.....	51
Chapter 4: Findings and Discussion.	53
4.1: Introduction.....	53
4.2: The sample and response rate.....	53
4.3: Participants’ years of experience and qualifications.	54
4.4: Teachers’ receptivity of the amount of information provided by the authorities.	55
4.5: Level of involvement in the reform.	56
4.6: The role of public perception in the reform.....	57
4.7: The influence of high-stakes international examinations.	58
4.8: Teachers’ workload.	59
4.9: Teaching and learning.	60
4.10: Issues of reliability, validity, and credibility.....	63
4.11: Teachers’ perceived roles, rapport, and accountability.	68
4.12: Teachers’ language testing and assessment training and professional development.	73
4.13: Teachers’ conceptualization of assessment.	74
4.14: Teachers’ self-efficacy and assessment literacy.	78
4.15: Conclusion.	86

Chapter 5: Concluding chapter.	87
5.2: Conclusions drawn from the study.	87
5.2.1: Secondary research question 1 (RQ2).	87
5.2.2: Secondary research question 2 (RQ3).	89
5.2.3: Secondary research question 3 (RQ4).	91
5.3: Limitations of the study.	91
5.4: Recommendations for policymakers.	92
5.5: Recommendations for future research.	93
5.6: Conclusion.	93
List of references.	95
Appendix A: Survey questions.	105
Appendix B: Interview questions. (Not used).	110
Appendix C: Permission letters.	111
Appendix D: Information letter. (Not used).	113

List of figures

<i>Figure 1: Participants' genders.</i>	53
<i>Figure 2: Participants' years of teaching experience.</i>	54
<i>Figure 3: Teachers' qualifications.</i>	55
<i>Figure 4: Teachers' perceived level of involvement in the reform.</i>	56
<i>Figure 5: Teachers' perceptions of the role of public perception in this reform.</i>	57
<i>Figure 6: Teachers' perception of the impact of the high-stakes examinations on the reform.</i>	58
<i>Figure 7: Teachers' perception of how this reform will affect their teaching methods.</i>	61
<i>Figure 8: Teachers' perception of how this reform will impact students' learning.</i>	62
<i>Figure 9: Teachers' confidence in their own ability to assess reliably.</i>	63
<i>Figure 10: Teachers' confidence in other teachers' ability to assess reliably.</i>	64
<i>Figure 11: Teachers' confidence in Teacher-Based Assessment as a valid way of assessing students' learning.</i>	65
<i>Figure 12: Teachers' perception on how the reform will affect the SEC qualification's credibility.</i>	66
<i>Figure 13: How teachers think the assessment tasks should be prepared.</i>	67
<i>Figure 14: Teachers' perception of how this reform will affect their relationships with students.</i>	69
<i>Figure 15: Teachers' perception of how this reform will affect their accountability as teachers.</i>	70
<i>Figure 16: Teachers' perception of how the general public will receive teacher-based assessment.</i>	71
<i>Figure 17: Teachers' perception of how this reform will affect their status as teachers.</i>	72
<i>Figure 18: Teachers' language testing and assessment training and professional development.</i>	73
<i>Figure 19: Teachers' conceptualization of assessment 1.</i>	74
<i>Figure 20: Teachers' conceptualization of assessment 2.</i>	75
<i>Figure 21: Teachers' conceptualization of assessment 3.</i>	76
<i>Figure 22: Teachers' conceptualization of assessment 4.</i>	77
<i>Figure 23: Teachers' self-efficacy when it comes to using assessment tasks to improve students' learning.</i>	79
<i>Figure 24: Teachers' self-efficacy when it comes to adapting their teaching as a result of students' results on assessment tasks.</i>	80
<i>Figure 25: Teachers' self-efficacy when it comes to preparing tasks in line with the SEC syllabus.</i>	81
<i>Figure 26: Teachers' self-efficacy when it comes to using rating scales consistently.</i>	82
<i>Figure 27: Teachers' self-efficacy when it comes to interpreting how reliable an assessment task is.</i>	83
<i>Figure 28: Teachers' self-efficacy when it comes to interpreting how valid an assessment task is.</i>	84
<i>Figure 29: Teachers' self-efficacy when it comes to understanding washback effects.</i>	85

Chapter 1: Introductory chapter.

1.1: The Maltese context.

One of the central features of all educational discourse is debate; indeed, debating different topics related to education allows various stakeholders to improve the quality of the educational provision, with this having many implications. One issue which is not debatable in the local context is the fact that our educational system is undergoing a radical overhaul in its set-up. Attard Tonna and Bugeja (2016) discuss the major reform which is currently being introduced in Maltese schools: The Learning Outcomes Framework. They describe this reform as being one which will shift the focus of Maltese education from “centrally-imposed knowledge-centric syllabi” to a “framework of knowledge, attitudes and skills-based outcomes that are considered the national educational entitlement of all learners in Malta.” (Attard Tonna and Bugeja, 2016, p.170). They further comment that this framework also entails a reform in the way in which student learning is to be assessed, with School-Based Assessment taking on a more leading position in a system which has traditionally favoured summative examinations. Given that this is a national reform, School-Based Assessment is being introduced at all levels of mandatory education, and across all curricular areas.

This dissertation focuses on the implication of this assessment reform in relation to Secondary School English. In order to understand this, however, one needs to understand the position of English in the local context; and, more specifically, the pedagogical and assessment traditions associated with this subject. The National Minimum Curriculum (hereafter, ‘NMC’) positions English as one of the core subjects, and stipulates that all students need to express “effective, precise, and confident use” of this language, along with the country’s national language: Maltese. (MEDE, 1999, p.23). Similarly, the more recent curriculum further confirms this as educational entitlement when it states that “young people in compulsory education [must] have as an indispensable prerequisite mastery in English and Maltese.” (MEDE, 2012, p.iii). It is hence clear to all involved in the Maltese educational system that learning English, and becoming proficient in it, is the entitlement of all young people.

Whilst both national curricula to have been published in the last twenty-one years ensure that English is the entitlement of all learners, the way in which this entitlement has been provided, as well as the assessment of whether or not it has been provided in a satisfactory manner, have faced numerous successes and challenges. A recent report by the *Council of Europe* discusses the general state of Language Education in the local context, by referring to both the teaching and assessment of English and other languages taught on the islands. One issue raised in the report focuses on the question of how language is assessed locally, and the effects this has on literacy and language learning in general. In this report, the fact that Maltese students do not generally fare as well as students in other countries on international tests such as PISA is highlighted. The report identifies Malta's high-stakes examination culture as being one of the plausible causes of this low performance by specifically stating that Malta has "a widespread culture which sets great store by benchmarks and testing", whilst then going on to comment on the notion of conducting High-Stakes Assessments at the age of sixteen, which it labels as being strange. (CoE, 2015, p.32). The report recommends significant changes to the system which is used to assess learners both throughout their educational journey, as well as at the end of it.

1.2: Statement of the problem.

In light of the *Council of Europe's* report, along with other factors which will be discussed at a later stage, a reform has been placed in motion which will change the way in which students will be assessed on their learning of both the English Language and English Literature at the end of their Secondary Education. Up until the year 2022, students will be assessed via a summative examination at the end of their schooling, for both subjects. (MATSEC, 2020, online). As of 2023, however, both subjects will include assessments in the form of coursework, which will be ongoing throughout the final three years of Secondary Education. (MATSEC, 2020, online). For English language, this is being distributed across the four language skills, with listening comprising 15% of the Continuous Assessment, and reading, writing, and speaking comprising 5% each. (MATSEC, 2020, p.26). The remaining 70% will then be assessed via a summative examination, as has been done traditionally, with listening to be

assessed solely through classroom assessment. For English Literature, students will carry out three assignments that test their ability to analyse and evaluate literary texts. The same ratio of coursework to examination will apply for English Literature. (MATSEC, 2020, pp.22-5).

Within the local context, this reform is radical. After many years of solely summative examinations, teachers are now required to take on a more direct role in their students' SEC assessment. Given the novelty of this approach in the school subject of English, this dissertation aims to explore how English teachers practising in the State Sector are responding to this change. Numerous factors may influence how teachers worldwide respond to educational reforms, and this dissertation will review the literature on several of these factors and then find out which ones are pertinent to the local context via the use of an online survey. The research design employs an essentially ontological epistemology in the hopes of shedding light on how teachers of English are responding to the assessment reform in the local context.

1.3: Frameworks considered.

This dissertation is influenced by three major frameworks which consider assessment from the perspective of the human experience: the socio-cultural approach to assessment, critical language testing, and teacher cognition. Up until the early 1990s, research on assessment was dominated by studies which focused on psychometric concepts, such as validity and reliability. (Spolsky, 2008). However, Spolsky (2008) points out that following this there was a transition in the way in which research on testing and assessment was carried out, with questions of ethics and social responsibility becoming more popular and, arguably, more necessary. Towards the end of his paper he highlights how testing is not an event which is done with neutral intentions, but rather, is positioned in the social sphere as an event which has implications for power relations. This is reinforced by the fact that test results, both locally and internationally, have a huge impact on the individual's life chances, be it when it comes to accessing higher education or the job market.

The concept being studied in this dissertation deals with an expansion in how assessment is being enacted in the local context, as what has traditionally been a summative

high-stakes examination is now being supplemented by assessment which is on-going in the classroom. Within such a context, it is necessary to consider how psychometric aspects, which according to Moss (1996) have always been the domain of the psychometric expert, will now come to be interpreted and used in a wider educational sphere consisting of classroom teachers. Teasdale and Leung (2000) consider the question of how well psychometric aspects can be tackled by teachers who have not had extensive training in psychometrics, and hence are not experts in the field of educational measurement. One of the central points which they raise is that “psychometric principles” cannot be exported from the expert’s table to the classroom in a manner which does not create problems, with two of these issues being “the stability of tasks” and “the conditions of assessment.” (Teasdale and Leung, 2000, p.167). Within the local context, these psychometric aspects are being exported to the classroom through this reform through how the new SEC syllabus requires teachers to be involved in the setting up and marking of tasks which will constitute a percentage of their students’ examination. To understand the implications of “the conditions of assessment”, one needs to divert one’s attention to the social sphere rather than the psychometric one, as this is where the conditions for such assessments are set.

Historically speaking, the notion of assessment originated in the social sphere and not in the psychometric one. Gipps (1994) points out that the first formal examinations ever held took place in China in approximately 206 B.C.E. The purpose of these examinations was to select people for the civil service, which highlights the social role which assessment plays in society. Gipps’ (1994) paper devotes itself to considering the socio-cultural approach to assessment and her main point is that whilst the psychometric aspect of assessment is valuable, this does not preclude the extensive social functions performed by and through assessment systems. In order to defend this thesis, she raises multiple points: how “examinations are a means of exerting power over others” (Gipps, 1994, p.366); how assessment in the postmodern age may be more of a reflection of the “assessor’s values” (Gipps, 1994, p.370); which assessment methods are best suited for minority groups (Gipps, 1994, p.376); and other social issues. This is pertinent to the topic being discussed in this dissertation as the social aspect of assessment has become more critical now that more people will be involved as assessors. The assessors’ attitudes and

values in relation to the functions of assessment, power, and maybe even minority groups, will now come to play a greater role in the outcome of this reform. Hence, it is sensible to study these people's receptivity of the reform.

The concept of language testing has been considerably criticised in the literature, with this also tying into social aspects of assessment. One such critic of the concept of the language test is Shohamy. (2001). In her book, the author discusses how at the heart of the language test one finds the notion of power to be a central theme. She points out that tests are "a method of control" (Shohamy, 2001, p.37) in how they allow the assessor to exert power over the tested person, and that this power is a result of the public's blind trust in such tests. Further to this, she claims that language tests attain their power as they are "administered by powerful institutions", "use the language of science", and "the language of numbers," amongst others. (Shohamy, 2001, p.21). Given that within the local context the shift is away from traditional testing, towards coursework in the form of written and spoken assignments in the SEC examination, Shohamy's critique becomes relevant in how factors such as power and the language used will affect the assessed. Whilst the teachers will now have more power, it is worth considering where the learner will stand.

The previous sections of this introduction have discussed the role that teachers will play in this reform, and this research aims to explore how teachers are responding to this reform. It is valuable, both in terms of academic research and in terms of policy development, to consider their receptivity as it is the teachers who will now have a more central role in the execution of these assessments. Research has shown that, when conducting assessments in class, teachers do not always follow the psychometric principles which educational measurement experts recommend. McMillan (2003:34) points out that various studies in different contexts have shown that teachers are more concerned with students mastering the content of a syllabus, than they are with measuring this mastery with precision. It would be inconsiderate to assume that this is simply because teachers are incapable of grasping the concepts associated with effective psychometric measurement, hence, there is the need for a deeper understanding of the cognitive processes underlying this behaviour. To understand this, the teacher cognition approach is germane. Teacher cognition is a diverse field, and this dissertation calls upon the

thinking of Woods (1994) who posits teacher cognition as being a blend of beliefs, assumptions, and knowledge (BAK) which teachers use to inform aspects of their practice. Woods points out that, traditionally, these three concepts have been “treated separately, but increasingly the distinctions between them are being blurred.” (Woods, 1994, p.185). The rationale behind this is that it is impossible to evaluate which practices are a result of beliefs, assumptions, or knowledge, specifically, and hence, most practice is a result of all three interacting with each other. Traditionally, the teacher cognition approach in language teaching has dealt with aspects such as teacher education, grammar, and literacy. (Borg, 2003). The researcher did not find any studies which approach assessment and testing from a teacher cognition approach, so this dissertation attempts to set a precedent in this aspect.

1.4: The Local Research Context.

Despite being a small country, Malta has a vibrant research community when it comes to education. Given that this dissertation focusses on teachers’ reception of a new assessment system, this section will contextualize the study in relation to a developing field of research on human responses to assessment in the local context.

The need for a reform in the state’s national examination system has been felt in the past, and both the state and the research community have been open to considering the perceptions of different stakeholders, including teachers. In a study commissioned by the state in 2003, Grima *et al.* found that coursework has always been a bone of contention in the Maltese context. They point out that in the twelve subjects which offered a school-based assessment component in the SEC examination, issues such as “low variance in marks awarded; inflated marks; differences between teachers’ marks and moderators’ marks” , amongst others, were found. (Grima *et al.*, 2005, p.11). This is controversial from an ethical perspective given that the implication of this is that students were not receiving the marks which they truly deserved. Further to this, the research also shows that the schools feel that coursework leaves “a positive impact on teaching and learning”. (Grima *et al.*, 2005, p.63). Whilst the schools, consisting of teachers amongst others, believed that the coursework had a positive impact on

the learning process, Grima *et al.* insist that the identified issues relating to a lack of reliability in these assessment components is of concern.

Local literature on coursework also reveals multiple other issues which teachers face when assessing parts of students' SEC examination. Rotin's (2004) study provides insights into these issues. Although the study focuses on Home Economics, and this dissertation focuses on English, it still provides valuable insights which arguably cut across subject-specific boundaries. For example, Rotin points out that one of the biggest challenges faced by teachers in the school was that there was not enough time for them to meet and discuss matters pertaining to assessment, due to the nature of their lesson schedules. Indeed, she points out that there was only one slot per week where teachers could meet to discuss. (Rotin, 2004). This is arguably detrimental to the process of assessing coursework as standardization could not be assured and therefore the reliability which can be achieved when two or more teachers work together was compromised. Other issues found in this local study consist of a lack of clear marking criteria against which to assess the coursework (Rotin, 2004); tasks done early on in the students' learning experiences but assessed against the same criteria as those done later on in process (Rotin, *ibid*); and students copying each other's work. (Rotin, *ibid*). Whilst this study is grounded in one school-subject, it would be premature to assume that these matters are unique to this subject. Through this study, one can see that teachers do face issues when course work is used to assess high-stakes examinations, and it would arguably be a mistake to dismiss these when researching such reforms.

A more recent study, this time with English teachers in Maltese Church Schools has also shown that teachers have different perceptions when it comes to assessment. (De Giovanni, 2019). In this study, De Giovanni conducted an online survey with teachers of English to understand their perception of assessment, as well as to consider what teachers think about the current SEC reform being undertaken. The study found that, within the Church Sector, English teachers see assessment as being "a means to measure progress", rather than a pedagogic tool which can be used to enhance the learning process. (De Giovanni, 2019, p.219). More pertinently, the study also found that these teachers "still favour national assessments to be carried out by MATSEC, despite the option that they could be carried out as school-based

assessments.” (De Giovanni, 2019, p.226). The author points out that this could be due to the reform not being completely in place yet, and highlights that this could require senior educational officials to provide more training for teachers. This study shows the significance of studying teacher perceptions in relation to assessment reforms, as it displays how further work may be needed by the educational authorities in order to avoid issues such as those identified in Grima *et al.* (2005) and Rotin (2004).

1.5: Research rationale and research questions.

In this section, the research questions for this dissertation will be identified and explicated. This study has one main research question and three secondary research questions which aim to examine how teachers in the State Sector are responding to having to assess parts of their students’ SEC examination via course work.

The main research question for this dissertation is:

RQ1: How are Maltese teachers of English responding to having to assess parts of their students’ SEC examination via course work, as of 2020?

This research question was developed after reading the work of Moroz and Waugh (2000) and Waugh and Godfrey (1993) whose studies both address the question of how teachers respond to “system-wide change” in given educational contexts. (Waugh and Godfrey, 1993, p.565). The latter point out that “it is important for educational decision-makers to know what variables affect receptivity” and, hence, this study aims to examine which factors affect how teachers are responding to this. (Waugh and Godfrey, 1993, p.565). If all goes well, it should be of interest to decision-makers who are involved with this reform.

The first secondary research question of this dissertation deals with different factors which affect teacher receptivity. The RQ is:

RQ2: Which factors/ variables have a strong relationship with English teachers’ receptivity of the local assessment reform?

For the purposes of this research question, an extensive literature review was conducted in which studies in different countries are considered. Once again, this RQ is influenced by the work of Moroz and Waugh (2000) and Waugh and Godfrey (1993) who point out that teachers' receptivity of a given reform is influenced by numerous factors. Their framework has been considered too generic for this research, however, so issues pertaining specifically to assessment reform have been identified in the literature review, with these then informing the data collection tools.

The second subsidiary research question for this dissertation deals with the effects which teachers think this reform will have on their working lives. The specific RQ is:

RQ3: How do teachers perceive that this change will affect their working lives, as teachers of English in the Maltese educational system?

This RQ is informed by the work of Fullan (2003) who points out that teachers are at the heart of any educational reform. Indeed, Fullan (2003) argues that very often "the purpose and passion which drives the best teachers" is often missing when educational reform is commenced. (Fullan, 2003, p.10). This question tries to consider whether the reform will have a change on the teachers' working lives, and if the change will be beneficial to them.

The final research question for this dissertation considers the relationship between English teachers' self-efficacy as assessors and their assessment literacy. It attempts to survey how efficacious the teachers are when it comes to assessing parts of their students' examination as part of their daily practice. The RQ is:

RQ4: How efficacious do teachers of English believe that they are when it comes to assessing the language skills and knowledge of their learners for the purposes of a High-Stakes Assessment?

Self-efficacy is being understood from the same perspective as Bandura (1997), as being the "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." (Bandura, 1997, p.3). The "course of action" in question here is English teachers' assessment literacy, which, in terms of language assessment, "refers to knowledge, skills, and principles in language testing." (Giraldo, 2018, p.180). These concepts are defined

with greater precision and delved into in more detail in the literature review of this dissertation.

Chapter 2: Literature review.

2.1: What is assessment reform?

2.1.1: Defining assessment and reform.

At the starting point of this literature review, it is worth pointing out that there are two assumptions embedded in this dissertation. The first one is that “educational assessment” is a dynamic field which, despite the efforts of many, evades concrete definitions. Newton (2007) points out that even amongst professionals who specialise in the area, there is “a lack of clarity” in the discourse surrounding this topic. (Newton, 2007, p.149) The inference being drawn from Newton’s article is that the term “assessment” evades definition as a concept when it is seen in isolation. Newton goes on to argue that, rather than quarrelling over definitions of assessment, a more worthwhile endeavour is to focus on the functions which assessment exercises execute, as the functions are what determine the consequences of any performed assessment exercise. Throughout this review, whenever the term “assessment” is mentioned, attention will be paid to the functions and the consequences of these functions, meaning that these will be seen within the context of various theories of assessment, policies, and practices. This dissertation does not seek to provide an all-encompassing definition or understanding of assessment as a construct in its own right.

A second assumption embedded in the work presented is that educational change is never straightforward and that mono-directional explanations of it are, due to the complexity of educational change and systematic changes in general, insufficient. This reflects the arguments made by Michael Fullan (1997, 2003) in his widely cited work on educational change. Fullan (1997) points out that all educational change takes place “on the edge of chaos” (Fullan, 1997, p.24), with this metaphorical edge being a location between excessive order and too little order. Fullan’s theory is posited within the context of managing change; however, it is worth extending this argument into the context of studying change. Any study on change will

be artificial if it attempts to pose too much or too little structure on the changes being observed. The “edge of chaos” hence plays a crucial role in the thoughts presented in the this review and explanations of change which are seemingly straightforward will be evaluated critically.

2.1.2: Reasons for studying reform in the English SEC examination.

This dissertation will be focusing on teachers of English for multiple reasons. The first reason is logistic: it would have been considerably difficult to study all teachers who are partaking in this change within the time constraints of the study. The second reason is personal: the researcher is in training to become an English teacher. This study will sensitise him to issues surrounding assessing English in class via continuous assessment. Also, the researcher is more knowledgeable in the English secondary school curriculum and the pedagogy which it entails than that of other subjects, implying that other subject areas would have fallen outside of the scope of his training to enter the teaching profession. It is also sensible to focus on English as a school-subject given that all students need to pass this examination to access higher education, as well as due to it often being a necessity for employment. This, in turn, warrants the usage of the terms “High-Stakes Examination” and “High-Stakes Assessment”, which are discussed later in this review.

2.2: Outline of the literature review.

With the assumptions and rationale of the target sample and subject discussed, the central arguments and outline of this review will now be discussed. The first part of this review will provide a critical review of the posited assessment reform within the Maltese context, and the argument will be made that the underlying scope of the reform is to enhance student performance in High-Stakes Examinations. The predecessor to this new system is hereby seen as providing unsatisfactory levels of student achievement, with this leading to pressure from various factions for a change in the way in which assessment takes place in Malta. These changes will then be analysed in relation to scholarly work on educational change in general, with this section being the point where the central argument of the review will be made. The central argument

being made is that teachers are at the heart of such an assessment reform and that their perspectives on the change are central to the outcome of this reform. Issues pertaining to Teacher-Based Assessment will also be tackled here, as they are also part of the problematisation of the given reform.

The second part of this review will focus on various factors, such as how teachers' knowledge and beliefs, cultural expectations, and others, have an influence on how teachers respond to assessment reforms. The argument posited here is that this system is complex in and of itself, and that there are multiple considerations which need to be taken into account prior to the execution of such a reform, due to how they can affect teachers' response to the new system. This complexity is further exacerbated when one considers that teachers of English will oversee assessing parts of a minimum of two SEC examinations, considering that English in Maltese state schools is split into two subjects: English Language and English Literature. This argument will then be further promoted by pointing out the need for further teacher training, or Professional Development (PD), in relation to the reform. Different models for this development will be reviewed and made relevant to the reform in the Maltese context.

The final part of this literature review deals with the interface between teachers' self-efficacy and assessment literacy. Assessment literacy is arguably a crucial component of this reform, given that teachers of English need to be literate both in terms of evaluating students' work, whilst also making the most of the assessment tasks to enhance student learning. Their self-efficacy in relation to assessment literacy is a critical point to consider, given that a component of their work will incorporate formal assessment.

2.3: A review of the reform.

2.3.1: The role of the public and culture in assessment reform.

In an informative 2001 article, Paul Black (2001:76) talks about assessment in terms of dreams and provides his readers with what the initial vision for formative assessment consisted of. Using the metaphor of a dream to describe the theoretical framework which Black has worked strenuously on throughout his academic career is an appropriate starting point for this part of the literature review. The word "dream" is appropriate because each type of

assessment presented by different scholars and policymakers who work in the field of assessment seem to be woken up, perhaps rather violently, by the fact that others do not share their vision. The fact that others might also have more sway than the initial “dreamer” is daunting for this person, but it also instils rethinking and, eventually, change.

The general public is worth being noted as one of the factions which can sway the way an initial idea pertaining to assessment, and education in general, can end up being changed. One of the central figures in the literature who focuses on this issue is Michael Fullan (2003), who writes that when the public criticises an aspect of education, this criticism needs to be taken on board as it is symptomatic of concern, rather than just being criticism for criticism’s sake. Fullan argues that when “the public is dissatisfied and anxious about the failure of the school system to keep up, ... this means they want to help fix it not abandon it.” (Fullan, 2003, p.17). More specifically, the literature on language teaching and language assessment has dealt with the issue of how public opinion regarding assessment influences such reforms. One such study is Mee’s (1998) paper, written in the Singaporean context, in which he argues that reforms made in language education need to “take into consideration examination-driven system[s]” and that any new approach to teaching or assessing needs to consider the context in which the said reform is taking place. (Mee, 1998, p.193). He further develops the concept of an “examination culture” by arguing that one can see this culture in effect by simply walking down the street and discussing the topic with natives of Singapore. The notion of public opinion having an influence on reforms in assessment is highlighted in this study through how the researcher argues that the examination system of the country is ingrained in its culture, implying that this, perhaps non-empirical, dimension of assessment plays a role in how reforms are executed and perceived by the general public.

Whilst Fullan’s work emphasises the crucial role which public opinion plays in educational reform, this does not mean that public opinion is absolute. He points out that educational change is often governed by a principle of auto-catalysis, which he cites from the work of Marion (1999, p.xii, as cited in Fullan, 2003, p.21). The concept of auto-catalysis is here being interpreted as being when particular behaviours in one part of a social system begin to change and cause changes in other parts of the system. This means that public opinion and

different factions in society, which include professional educators and educational experts, interact with one another in order to bring about changes within the system, with the interaction being spearheaded by one group, and resulting in changes in the others, which in turn change the whole system. In this case, the public is a social organism which is plausibly changing its attitude towards the way in which assessment is conducted in Malta, and educational policymakers who deal with curriculum and assessment are responding by changing the system. In a comparative study, Davison (2004) points out how a reform remarkably like the Maltese one, consisting of Teacher-Based Assessments as part of a High-Stakes Examination, was received contrastingly in different cultures. Davison compares the receptivity of the same assessment procedures in Melbourne, Australia with that of Hong Kong. Prior to collecting data, Davison (2004:313) expected the teachers in Australia to mark students' work with greater accuracy than those practising in Hong Kong given that they had much more experience in marking coursework as part of high-stakes examinations; however, the hypothesis turned out to be untrue. The explanation for this lies in how the teachers in Australia were more inclined to use their professional judgement, due to there being a strong belief in this judgment by the public, whilst in Hong Kong the public placed most of their trust in the examination system. Essentially, the latter led to the teachers wanting to approximate their marking strategies to those used by the examination system, rather than use their own professional judgment to assign marks. This shows the principle of auto-catalysis in action in how public attitudes towards how assessment should be conducted were an influence on how the teachers responded to the task set for them by the educational authorities, which then in turn influenced the grades which the students received from the teachers. Whilst in these contexts the reforms were not initiated by the public, it is clear that public perception plays a role in influencing the behaviours which come after the enactment of a reform.

2.3.2: The pressure for reform and the Maltese response to this.

The need for a reform in the way in which High-Stakes Examinations are conducted has been a recurring theme in the local context for many years. In the NCF, one notes that the authors implicitly criticise the notion of using examinations as a sole means of assessing

learners' summative performance when they write that "assessment of the students' achievement for summative purposes should not be dependent on one-off performances in tests and examinations." (MEDE, 2012, p.81). This taunts the previous system used at SEC level for English, as well as other subjects, when one considers that these subjects had always been assessed via examinations at the end of Secondary Education prior to this reform. Naturally, this kind of criticism of the examination is far from new. For example, Black (1998) cites records from the nineteenth century in which a Cambridge don is seen as criticising the way in which exams can be determinants of one's future. This section of the literature review deals with various factors which are arguably fuelling the urge for an assessment reform in Malta.

One factor which has caused considerable debate in the local context is Malta's below-average scores on international tests, such as PISA and PIRLS. In the 2015 PISA assessments, Malta scored 465 on average, compared to the global average of 493. Whilst these scores do not place Malta in the lowest scoring percentile, the fact that the country is still scoring at a below-average rate is of concern. (OECD, 2018). Similarly, in PIRLS, Malta's score was significantly below the global average in reading: where the average student around the globe scored 500 points, the average Maltese student scored 452 (MEDE, 2016, p.19). Whilst Mifsud (Mifsud for MEDE, 2016, p.vii) points out that this underperformance is attributable to the learners being assessed solely in the Maltese language when a number of students in the country are in fact English-dominant, the point still stands that we are comparatively lagging behind other countries. Further to this whilst PISA claims to separate content from literacy, that argument is somewhat difficult to accept given that understanding the content of what one reads is inextricably linked to one's ability to read (Goldstein, 2004). With this being said, one cannot ignore the fact that Malta's scores show a country which is lagging behind when compared to other countries, and whilst it is possible to explain this from a linguistic aspect, the point still stands that this affects the image of the country. To improve this, reform is necessary in terms of both pedagogic practice and assessment practice.

The need to change pedagogic practice, and assessment practice with it, can further be explained via the work of Fullan (2003) who asserts that for educational change to be executed successfully, the starting point is "to change the immediate context"(Fullan, 2003, p.27). Fullan

calls upon Gladwell's (Gladwell, 2000 as cited in Fullan, 2003, p.27) Broken Windows Theory, which is a criminological theory that asserts that when a neighbourhood appears to be run down, the chances of crimes taking place are much higher. "Fixing the windows" and maintaining the upkeep of a given neighbourhood will hence result in less crime. Fullan's application of this to general change in education stipulates that for matters to change for the better, changes must occur within the immediate context of stakeholders' lives which attempt to improve the factor which needs to be improved. Not attempting to improve whichever factors need improving will inevitably result in it deteriorating rather than improving. An application of this theory to the current context of achievement in Malta would argue that: if there are no serious attempts at improving students' achievement, there is a decent probability that attainment levels will plummet, rather than increase. This therefore means that for educational achievement to be increased, reforms will be needed which directly aim at increasing student achievement.

In order to increase achievement, the Maltese government has in fact embarked on a series of reforms. The broad scopes of these reforms can be seen in the *Framework for the Education Strategy for Malta 2014-2024*, which specifically lists increasing student achievement as one of its top priorities (MEDE, 2014). Amongst these reforms, one finds two pertinent ones which pertain to both the teaching and assessment of all subjects, including English. Spiteri (2019), writing about this transition within the context of the English syllabus, describes it as placing "learning and the learners at the heart of its operations". ; however, she also argues that the assessment reform followed too quickly the teaching reform ensuing from the Learning Outcomes Framework. (Spiteri, 2019, p.195). In terms of assessment, the *Learning Outcomes Framework*, pays specific attention to the use of strategies which go beyond examinations or summative assessments and argues that both the English Language and English Literature SEC qualifications should be assessed via a blended approach.(Directorate for Standards and Quality in Education, 2015). This reform is in its second year of the implementation stage, with ongoing assessment, which, in theory, can be both formative and summative assessment given that the syllabus calls for both types to be utilized to assess learners.

The second reform in question deals with a change in how the students' achievement, as a result of the reform in the students' learning, will be measured. For the purposes of this review, the proposed reform will be analysed through the lens of the subjects which fall under the responsibilities of English teachers – namely language and literature. The English language exam is being interpreted here as being a High-Stakes Examination, as students who wish to follow courses which lead to higher education or who wish to attain skilled or semi-skilled employment, are required to pass this examination. The English literature examination, on the other hand, does not possess the stakes of the language exam as a pass grade is not mandatory for one to proceed to higher education. The proposed change is simple to explain, yet complex in its implications. In its current state, the English SEC is assessed solely through summative means, organized by MATSEC, an examination board which is external to secondary schools in Malta and Gozo. With the proposed reform, the SEC certificate will continue to be partially assessed via a final examination; however, a certain percentage will be assessed by teachers, via “School Based Assessment in the form of continuous assessment.” (Directorate for Learning and Assessment Programmes, 2018: online). The implications of this will now be considered.

2.3.3: The political implications of the proposed reform.

One of the major implications of the proposed assessment reform at SEC level is that teachers are now being given the role of game-changers in Malta's achievement scenario. Whilst the official discourse on the topic is revolving around schools and classrooms, it is worthwhile to remember that teachers are at the helm of these classrooms. The notion of expecting teachers to change their practices, with specific reference to assessment in the classroom, is in line with a strand of assessment theory which attests that increasing achievement needs to start in the classroom. Indeed, in a text which has now become a classic in the assessment literature, Paul Black and Dylan William (2010) point out that “policy ought to start with a recognition that the prime locus of raising standards is the classroom, so that the overarching priority has to be the promotion and support of change within the classroom.” (Black and William, 2010, p.88). Their comment is written within the context of their argument

that formative assessment, conducted by teachers, is one of the key ways to improve learners' learning.

The notion of creating reforms which further places teachers at the centre of increasing learner achievement is also relevant with the context of local attitudes towards teachers and education. In a study on lifelong learning in Malta, Borg *et al.* (2016) found that "most of the participants see teachers as the most important influence in education." (Borg *et al.*, 2016, p.60) The survey also shows promising data on the general population's belief in teachers' abilities to motivate learners, as well as their subject and pedagogic knowledge. Whilst Borg *et al.*'s study does not deal with the topic of educational assessment and its reform in Malta, it is worth pointing out that this data easily lends itself to a justification, from the state's side, of entrusting teachers to be central figures in this assessment reform.

The state's choice to give teachers more responsibilities when it comes to the assessment of students' SEC examination is reflective of the adoption of particular socio-political perspectives. One of the central figures to criticise the social functions of the examination as an assessment technique was the French philosopher, Michel Foucault. In *Discipline and Punish*, Foucault (1977) writes that the examination changes "the economy of visibility into the exercise of power." (Foucault, 1977, p.60). Essentially, this means that exams are an economical way for the powerful to be able to see, evaluate, and eventually control examinees who eventually become the subjects of the powerful. The element of subjectivation is seen in the nature of the examination as an impersonal assessment technique, evident in most educational contexts through the anonymising of examinees' scripts. This process of subjectivation, within the context of the given change, is being subverted via continuous School-Based Assessment in that teachers are now being given the authority to assess parts of their students' SEC examinations. Whilst the purpose of this argument is not to imply that the reform is Foucauldian in nature, it probably is not, Foucault (1977:184) would argue that the "normalizing gaze" has penetrated even further as there are more observers in the panopticon now, considering that every teacher now has at least one finger in an examiner's pie. (Foucault, 1977, p.184) Foucauldian insights do allow for the point to stand that teachers are now being

entrusted with further responsibilities, and this reflects a socio-political perspective which sees teachers as being trustworthy of this role.

The previous paragraph acknowledges the Politics, with a capital “P”, of the proposed assessment reform; however, it is worth considering that the implications of placing teachers in this new position of authority from the perspective of politics, with a small “p”. This is the case as the reform in the Maltese context shares many similarities with similar reforms in the 1990s and early 2000s within the British educational system. Firestone (1998) points out that the early 1990s in the United Kingdom were characterized by assessment reforms which run contrary to the reforms happening in Malta, whereby examinations were being given further emphasis to the detriment of teacher-based assessment. Firestone (1998) characterises this as being the Conservative Party’s reaction to “too many trendy subjects, too much progressive teaching, and a drift away from ‘real testing’ and exams.” (Firestone, 1998, p.179) Similarly, Whetton (2009: accounts for this by pointing out that these reforms were due to a “Thatcherite distrust of established professional interests”, with the clear implications of both Firestone and Whetton’s comments being that the Conservative administration of the time did not believe in teachers’ capacity to conduct assessments, and that this task was to be entrusted to government-appointed measurement experts. (Whetton, 2009, p.138) Later down the line, however, a Labour government led by Tony Blair would introduce a system in which Teacher-Based Assessment was introduced to stand alongside examinations and criterion-referenced testing (Whetton, 2009). This shows that the issues being encountered in Malta are not, in any sense, unique. This literature review will now move on to consider issues encountered when similar reforms were enacted in different countries, in order to examine possible issues which the Maltese reform may encounter.

2.3.4: The Influence of Formative Assessment.

Two researchers who are of significance to the study of classroom assessment are Paul Black and Dylan William, scholars who are commonly associated with the advancement of classroom assessment in the form of formative assessment. Formative assessment is loosely defined by Black and William (1998a) as

“...encompassing all those activities undertaken by teachers, and/or their students which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.” (Black and William, 1998a, pp.7-8)

Black and William (1998a) draw attention to significant evidence that shows that tenets associated with formative assessment, namely timely feedback, are associated with increases in student learning. The Maltese state’s proposed reform is tentatively in line with this theoretical persuasion on assessment as, through increasing classroom time and introducing mandatory School-Based Assessment, these practices are being viewed as springboards to greater academic achievement. Of course, this is not to say that nation-wide summative assessments, in the form of examinations within the local context, are futile or expendable. As Black (2001) suggests:

“Because anyone with educated capabilities requires proof of their achievements, summative individual assessment will always be essential, but the effort required for its attainment may strain any affection with the process.”
(Black, 2001, p.154)

This reform aims to encompass both the advantages of formative assessment, as well as retain the necessary academic and social functions associated with summative assessment.

Whilst formative assessment and summative assessment have their respective advantages when looked at from multiple perspectives, intertwining the two in the form of Continuous School-Based Assessment, as the current reform proposes, has not occurred without a set of issues which are unique to these types of assessment policies. A factor which cannot be dismissed is that teachers are at the heart of this reform. They are in such a position as they will have to take on the responsibility of assessing parts of their students’ SEC examinations. In the case of the SEC for English language, the stakes are much higher considering that passing this assessment is mandatory for learners to advance to post-secondary education and find work opportunities. Various issues which arise when assessing parts of high-stakes examinations in class will now be considered.

2.3.5: Factors affecting the successful implementation of the reform.

2.3.5.1: Teachers' Conceptualization of Assessment

The way in which teachers conceive of this reform is crucial to its successful implementation. Brown (2004:301) points out that the success, or failure, of assessment policies all depends on how teachers conceptualise any given assessment policy. He then goes on to argue that their conception of assessment policy must be understood against three criteria: how it is understood, how it is responded to, and how the teachers opt to change their interactions with other stakeholders in the educational sector as a result of it. In his study with New Zealand Secondary School teachers, Brown found that factors such as gender, experience, and amount of training shared no significant relationship with the way in which teachers conceptualize the concept. However, his study does show how the teachers all had diverse ideas on what assessment is, what its purpose is, and how it should be carried out. The significance of this study lies in how assessment policymakers need to be explicit when it comes to the purpose of assessment reform, as well as in what they expect from teachers as part of the reform.

The success of our reform depends on another crucial point: Continuous Assessment and Assessment for Learning are not the same concept (Harlen and James, 1997). In itself, Continuous Assessment is understood by Harlen and James (*ibid.*) as assessment which is ongoing, with this “ongoing-ness” not necessarily consisting of solely formative or summative assessment, and neither consisting of a mixture of the both of them. Indeed, it is difficult to pin down what the correct practice associated with Continuous Assessment should be, as it is a vague term when it is not defined within a specific context of practice. This, in turn, has implications for teachers' practice, as it is the teachers who must come to terms with the idea that the assessments which they will have to conduct must integrate formative aspects, due to these being held in the classroom, with summative aspects, because of the SEC.

2.3.5.2: Issues of Reliability and Validity.

One of the implications of this is that teachers will have to ensure that the assessments conducted in class for the purposes of the SEC examinations are valid and reliable. Reliability is interpreted by Black (2001) as occurring when the student achieves the same result for the same learning outcome in many different contexts, meaning that the same student should achieve the same result for his achievement on comparable measures. Black (2001) also defines validity as “whether a test will measure what those who prepared it intended to measure.” (Black, 2001, p.42) Whilst in this case the discussion centres around assessment in its broadest term, and is not specific to tests as there is no mention of Continuous Assessment being a series of tests in the classroom in the English syllabi, these two concepts are still crucial from the perspective of how teachers will respond to the given reform. It is highly likely, however, that testing and other summative assessment techniques will dominate this scene, given that teachers will inevitably be required to produce a result which will complement the students’ final examination, and formative techniques do not produce quantitative data which can be used for these purposes (Black and William, 1998a). In several contexts, these two factors have played a considerable role in how teachers were perceived to be doing their jobs by other stakeholders, with this inherently influencing teachers’ receptivity of the given reform.

Questions pertaining to validity and reliability are questions of fairness, and when these are not present in assessment events, these are not fair assessments. This, in turn, should motivate teachers to create assessment tasks which are reliable and valid; however, this may not always follow. Consequentially, this can result in guilt on the teachers’ sides, with guilt being “a central emotional preoccupation for teachers.” (Hargreaves, 1994, p.142). One study in the UK context has yielded data which is well-worth considering in terms of how teachers deal with the questions of reliability and validity. Wyatt *et al.* (2010) argue that teacher-based assessments tend to be “viewed as having high validity but questionable reliability.” (Wyatt *et al.*, 2010, p.61). Similarly, they also point out that, even if an authoritative body were to provide

teachers in different schools with the same assessment material, different results will also be yielded due to biases. (Wyatt *et al.*, 2010). Reliability is also further compromised by the fact that teachers also have their own conceptions of what constitutes a proficient learner, with this being reflected in how teachers assess students' work. (Wyatt *et al.*, 2010). Whilst the authors contend that reliability is a bigger issue than validity, it is worth considering that validity, too, should be a central preoccupation. This can only be achieved through task-designers being in possession of the right assessment literacy to create assessment tasks which assess that which they are meant to assess. The issue of teachers' assessment literacy is delved into in more depth in the final part of this literature review.

Within the context of English in secondary education, Teacher-Based Assessment in High-Stakes Assessments yielded undesirable results in the UK. Isaacs (2014) uses the geographical metaphor of a "perfect storm" to describe how this type of assessment policy panned out, due to the chaos that ensued. Whilst Isaacs highlights several factors which caused a public outcry over the assessments, one of the key issues identified was that English teachers were conflating marks. Isaacs (2014) points out that:

"Consciously or subconsciously, teachers want as many students as possible to obtain that iconic grade and unlike GCSE Mathematics, English is an open-ended essay-based subject where the best marking scheme in the world cannot (and should not) exactly describe the entirety of student performance." (Isaacs, 2014, p.39).

This issue is not just unique to English assessments. Research has also shown that other factors which are not related to students' academic achievement which influence the grades assigned by teachers include "attendance, effort, disruptive behaviour, and completing work assignments." (Willingham, Pollack, and Lewis, 2002, p.10). This is a rather worrying factor which will need to be considered, and an understanding of how teachers are responding to the reform will arguably allow for a better understanding of which factors are at the top of their concerns.

The research of Wynne Harlen (2005a, 2005b) is also useful in further expanding the argument that teachers are at the heart of such reforms, and how reliability and validity in assessment are tied into this. Harlen (2005b) points out that:

“...if validity is increased by extending the range of assessment to include outcomes such as higher-level thinking skills, then reliability is likely to fall, since many of these aspects of attainment are not easily assessed.” (Harlen, 2005b, p.247)

The implication of this is that teachers, both individually and collectively within a given school, will have to decide as to whether reliability or validity will take priority. She points out that uniformity can only be achieved in classroom-based assessments “if the classroom is turned into an examination room.” (Harlen, 2005b, p.265). This all ties in together when one considers that the proposed reform will essentially equate to different students in different part of the Maltese islands doing different assessment tasks for the purposes of the same qualification. Should one teacher, hypothetically, attempt to prioritise reliability whilst another teacher prioritises validity, then one can conclude that students are being assessed via different standards, and therefore the assessment is not fair. It is the teachers who will be at the heart of this choice, considering that the assessments will be taking place in their classrooms; and hence, it is up to them to consider the issues of reliability and validity. Similarly, Harlen (2005a) also points out that one of the central disadvantages of Teacher-Based summative assessments is that there is “evidence of unreliability”, as well as the requirement of further resources in the form of moderation to circumvent this. (Harlen, 2005a, p.212) Whilst moderation will be necessary, it is highly unlikely that moderators will be able to moderate every single classroom assessment within an entire country. The reality remains that teachers will remain responsible for ensuring validity and reliability within their assessments, and this cannot be guaranteed.

2.3.5.3: The washback effect.

Washback and assessment often go together, with the former having a strong influence on the latter, and vice versa. Alderson and Wall (1992) point out that there are, broadly

speaking, two types of washback. Negative washback refers to when an assessment, usually a test or examination, has a negative effect on the teaching and learning process. For example, this can refer to when rote learning replaces meaningful learning for students to be able to score highly on a test or exam. Hamp-Lyons (1998) argues that within the context of English Language Teaching, high-stakes tests can have detrimental effects on the learning process, given that tests can only attempt to replicate real-life language use. On the other hand, positive washback is seen as a result of a well-constructed assessment. Alderson and Wall stipulate that if a test is good, then it should have a positive effect on learning. This does, however, have its limitations as the relationship between testing and learning is not as straightforward. Research shows that the relationship between testing and learning is based on a principle of diminishing returns: increasing the frequency of valid tests will only increase learning up to a certain degree. After this, testing has no effect or a detrimental effect on learning. (Bangert-Drowns *et al.*, 1991).

Washback is also seen in how summative examinations can come to harm the formative value which may be allotted to given assessment tasks. (Black *et al.*, 2010). Black *et al.* (2010) point out that there is a tendency for students, parents, and other stakeholders to value results from summative assessment more than that allocated to the ongoing assessments. They also argue that such classroom assessments are futile if they are not taken into consideration and exploited for the purposes of learning. (Black *et al.*, 2010). This is problematic, especially for teachers, as the reform requires teachers to conduct assessment tasks which are part of the final result, yet their efforts will be in vain should they not be exploited for learning purposes. Exploiting the tasks for learning purposes will, however, compromise the finality associated with summative examinations in how the assessment will not be summative, albeit contributing to a summative result. This is a dilemma which teachers will have to resolve in their classroom, which further shows that teachers are at the heart of this assessment reform.

The way in which teachers will resolve the various issues raised further shows that it is the teachers who are at the crux of this reform. Their decisions will arguably determine the outcome of this system. Factors within the system which can determine how teachers respond to this reform will now be considered.

2.4: Factors and variables determining teachers' receptivity.

The previous section of this literature review established that teachers are at the centre of the Maltese assessment reform. In this section, issues with understanding how teachers respond to given reforms will be considered. The underlying rationale for delving into different studies on the variables which affect teacher receptivity is to explicate the factors which typically correlate with teachers' receptivity of assessment reforms.

2.4.1: A definition of 'receptivity'.

Prior to considering the issues which relate to how teachers in different contexts have responded to such reforms, one needs to define what "receptivity" means. In this research, the word "receptivity" is to be understood as being semantically related to literature which discusses beliefs teachers hold about many aspects of their practice. Pajares (1992) points out that it is difficult to distinguish between what is considered to be knowledge and belief within the teaching profession in general, due to the interconnections between what is considered to be factual and research-based, and what the individual considers to have been learned through experience by one's self. Within the context of assessment, this corresponds to the point by Harlen (2005b) in terms of how teachers respond to validity and reliability, yet it also extends beyond this. Several beliefs or views held by teachers, of different interconnected aspects, can lead to differing levels of receptivity of the given assessment reform. This is in line with Pajares' (1992) argument that beliefs must be seen in context, as in this case there are two contexts: the context of assessment reform, specifically, and the context of the Maltese educational system, more generally, which can have a bearing on how teachers respond.

2.4.2: Teachers' knowledge and beliefs.

The notion of studying what teachers know, think, and feel about teaching is far from new. In the 1990s, however, a framework known as the Teacher Cognition approach began to formalize itself as a research framework which focuses on the actions which teachers engage in in class, as well as the thought processes, assumptions, and beliefs behind it. One of the key

figures in this approach, Woods, points out that whilst syllabus-designers and policymakers create the documents which outline what is meant to happen in the classroom, “the teacher usually plays a crucial role in determining what happens, on a moment-to-moment basis.” (Woods, 1994, p.13). Due to this convincing argument, it is sensible to discuss not only how teachers conceptualize assessment, but also what they believe to be appropriate in terms of assessing learners. The reason for doing this also lies in how a change in behaviour does not necessarily imply a change in what the teachers believe to be what their students need or deserve. Due to this, Borg (2006) points out that it is important to directly ask teachers about given aspects of their practice. He argues that:

“Teachers may adopt and display particular behaviours without any accompanying change in their cognitions (e.g. during assessed teaching practice, visits by inspectors or observations by researchers); teachers’ cognitions (e.g. their belief in the value of a particular instructional technique) may also change without any obvious change in what they do (due to, for example, situational constraints).” (Borg, 2006, p.277).

The implication for this to the assessment reform in Malta is that the teachers will be obliged to record students’ progress for the purposes of their SEC assessment; however, this does not mean that the thinking behind teachers’ practices in the classroom are the same, nor that they follow the thinking of those that developed the reform.

Whilst policy documents are usually written to attract, it follows that teachers and other educational stakeholders may not always follow in the belief that the reform is beneficial. van den Berg (2002) argues that what constitutes worthwhile change by policymakers will not always correspond with what teachers consider to be “good teaching”. Indeed, van den Berg (2002) maintains that the meanings which teachers ascribe to their practice, which he calls their “personal subjective educational theory”, is a type of knowledge which the teacher develops through experience. Similarly, teachers also develop this type of thinking regarding assessment, with this being reflected in various attitudes towards examinations, testing, and other assessment types. One of the final, relevant points which van den Berg (2002) makes is that

teachers have their own unique view of themselves as teachers. When teachers are expected to change their role, and this assessment reform is asking them to change their role in how they are now also assessors of their students' SEC examination, feelings of role ambiguity can develop, and this can eventually lead to stress in van den Berg's perspective.

The same assessment task may be defined and understood differently by different teachers, with this being an essential starting point in how the ascribed definition which teachers allocate to a given assessment task may become a confounding variable in their attitude towards it. This means that the definition ascribed by teachers to Continuous School-Based Assessment may not be the same. This issue was studied by De Lisle (2015), within the context of Trinidad and Tobago. In this study, De Lisle analysed how teachers conceive of continuous assessment, and correlated it with the construct of formative assessment. His research found that the factors which influence teachers' conceptions of Formative Assessment and Continuous Assessment are significantly different, with their being a weak correlation between the two. In the final part of his paper, he writes that

“Both the qualitative and quantitative findings suggested that even when teachers adopted CA, they ignored the formative intent desired by the programme planners.” (De Lisle, 2015,.97).

The inference drawn from this is that teachers can undermine the purposes of given assessment reforms and might even be doing this unintentionally through not understanding the rationale behind the reform, or not through not believing in it.

Teacher's personal theory of teaching may also have an impact on how reforms are perceived by teachers, and this then influences the behaviours which will follow. The question of how assessment reforms can be perceived differently is also raised by Hargreaves, Earl, and Schmidt (2002). Whilst their reflections centre around Grade 7 and 8 students in Ontario, Canada, it still provides valuable insights into how the same assessment concept can be seen differently by practitioners in the same context. Some differences between this system and the proposed Maltese one is that, within the Canadian context, final examinations were removed, and assessment was left entirely to the teacher. (Hargreaves, Earl, and Schmidt, 2002). In their

reflections upon this, they analyse the given reform from multiple theoretical vantage points. They point out that the concept of classroom-based assessment can be seen as humanistic in how it focuses on student-support; however, they also point out that it can be a means of surveillance in how teachers are constantly observing their learners' grades and marks, with few gaps between one assessment and another. (Hargreaves, Earl, and Schmidt, 2002). What Hargreaves *et al.* (2002) do not mention, however, is that teachers may also take sides in this debate, even if not consciously. Essentially, this depends upon what teachers believe about the reform as a result of their experience and knowledge.

Later research on how teachers conceive of assessment reforms has also been connected to how teachers believe that these reforms will affect their students' learning. (Brown *et al.*, 2009;2010). Their research was conducted within the contexts of New Zealand and Queensland, respectively. Their 2009 study focuses on how teachers opt to change their teaching methods in relation to reforms in assessment policies. Within the context of New Zealand, the teachers are entrusted with assessing parts of their students' final examinations at the end of secondary school, which makes the policy very similar to the proposed reform in the Maltese context. In New Zealand, the researchers found that the teachers "slightly agreed with the transmission perspective" of teaching, and they later explain that part of the reason for this is that they hold themselves, and are held by others to be, more accountable for the results which students attain in this high-stakes assessment. (Brown, *et al.*, 2009, p.69). In their 2010 research, with a different sample group in the same country, these researchers yielded further thought-provoking findings. Their study shows that the teachers were not anti school-based assessment, but rather, saw the assessments as being accountability-measuring methods for themselves, which they then imposed as learning-accountability measures on their learners. The conclusion drawn from the research is that teachers were only willing to put in an extensive effort to carry out their duties related to school-based assessment if they saw the task as being worthwhile and beneficial for learning. (Brown *et al.*, 2010). The implications of these two studies are that when assessments are used as a means of holding teachers accountable, rather than to improve student learning, then the attitude tends to be negative. The correlation presented between direct instruction and school-based assessment at the secondary school

level is of concern, considering the implications which this can have on the formative aspect of assessment. In fact, Hattie (2012) points out that, out of one-hundred-and-fifty possible variables which have been proven to have an influence on student achievement, “providing formative evaluation” and “feedback” are in the top ten. (Hattie, 2012, p.266). These aspects are often missing in direct instruction, which often implies a lecture-style of teaching.

2.4.3: Cultural expectations and their impact on assessment reform receptivity.

One of the central factors which may determine how teachers respond to assessment reforms is the culture in which the teacher practises. In two articles based on the Hong Kong educational system, Cheong (1999) and Berry (2011) focus on the question of how culture affects a society’s chosen way of assessing learning, as well as how a society looks at those who are responsible for this assessment. In his reflections, Cheong (1999) points out that even within Hong Kong’s heavy examination culture, there was still a School-Based Assessment system which fed into students’ high-stakes examinations. He does point out that, within such a culture, there is a need for teachers to be supported by various stakeholders. Given the importance attached to exam results in Hong Kong, it is sensible to argue that the teachers would require more support as the pressure and expectations placed on them are greater. Later research by Berry (2011:201) points out that within the timeframe in which Cheong was writing, this type of assessment took on the role of relentless testing which centred itself around tasks similar to those in the final examinations which learners would take. In 2011, Berry also points out that embracing school-based assessment to promote formative assessment (and hence, learning) ended up resulting in more frequent testing. Formative classroom assessment was hence compromised by the value placed on summative testing and examinations. The issue of how important a culture thinks an exam is shown to have an impact on what practices are conducted in class, as well as the pressure which teachers feel as a result of these cultural expectations. Hence, the culture in which a teacher practises, as well as their perceived position within this culture, will determine parts of their receptivity of this system.

A further issue which can determine how teachers respond to assessment reform is that of how teachers respond to the authority of a governing body. This factor intertwines with the

previously mentioned factors through how it encompasses both the teachers' individual theoretical disposition towards the given reform, as well as that of a culture's attitude towards authority. This factor was a determining variable in a study by Nsibande and Modiba (2012). In this study, the assessment reform which introduced school-based continuous assessment was seen as an imposition from above, with teachers just doing the bare minimum amount of work required to satisfy their superiors. The notions of coping and the assessment tasks taking too much time out of the teachers' day were also raised by the participants in this study. The issue being extrapolated from Nsibande and Modiba's work is that different teachers in different contexts have different attitudes towards authority. Within their context, the sample seems to be complying with the reforms, despite admitting to not putting in an extensive amount of effort into the assessment tasks which they needed to use to assess their learners. This shows how teachers' attitudes towards authority can have an impact on their receptivity of an assessment reform.

The historical and political circumstances under which teachers practise may also have an impact on how they view and respond to different assessment systems. Tavcar Krajnc (2006) documents the introduction of school-based assessment as being a means of resistance against the former communist philosophy of Slovenia. Tavcar Krajnc (2006) writes the following, in relation to a proposed reform to remove school-based assessment in favour of more extensive nation-wide examination-based assessment:

“The experts as well as the general public resisted the reintroduction, while the strongest reaction came from the students themselves, who in 1991 organized public protests against the introduction of a systematized secondary school completion. Consequently, the final examination was introduced in a limited form, with a smaller number of exams (two subjects instead of the planned four) and with an internal system of assessment carried out by the schools themselves.”
(Tavcar Krajnc, 2006, p.123)

Whilst no new constructs can be drawn from this, it further enforces the argument that teachers' reception of given reforms can have a significant influence on how successful a

reform can be, with this being at least partially determined by their culture and attitudes towards authority.

The effect of culture on the receptivity of assessment policies has also been seen to be connected to social values held within a culture. This was revealed in a study with English teachers in Bhutan, where School-Based Assessment is also used to measure a part of students' final examination at the end of mandatory education. (Luyten and Dolkar, 2010). Luyten and Dolkar's research was a response to the substantially lower results students attained in their final examination compared to the School-Based Assessment tasks which their teachers marked. They conclude that the explanation for this is that the teachers were using assessments which were not in line with the examination, despite the assessments supposedly measuring the same concepts which were to be assessed in the examination. (Luyten and Dolkar, 2010). As is the case in Isaacs' (2014) study, the marks given in the school-based assessment were inflated when compared to those achieved in the final examination. This may be explained culturally in how Bhutan measures national success in a different way to other countries: in this country, national success is measured via a Gross Happiness Index rather than through the well-known Gross Domestic Product. (Luyten and Dolkar, 2010:). Whilst Luyten and Dolkar do not draw conclusions based on this, they do point out that the rationale behind decentralizing assessment lies in this, as there is a strong belief in making students and the citizens of the nation happy. In this society, happiness is prioritized over wealth, with the former being rather difficult to measure and the latter being measurable with military precision. Hence, it is worth considering that this perception could easily blend in with the way in which teachers mark their students' work and result in them awarding students with results which they did not actually attain in order to retain their happiness.

The way in which teachers mark assessment tasks varies depending on the cultural context in which the teacher works. Within the Namibian context, Samson and Marongwe (2013) studied the distinctions between teachers' marks and the marks awarded by external examiners. This study was conducted with teachers of Mathematics; however, its findings are still arguably relevant to a study with English teachers in how the explanations provided by the teachers are not subject-specific. The starting explanation lies in how 30% of the teachers saw

the School-Based Assessments as being summative, rather than formative, in nature. (Samson and Marongwe, 2013). This implies that they saw the assessments as needing to be conducted under examination conditions. The rationale behind why they were stricter when marking is, however, rooted in social perceptions of the tasks at hand: they felt that the assessments were seen as invalid considering that they had less weighting than the examination, meaning that the examination was more important to them. This is in line with the previous discussion on the work of Cheong (1999) and Berry (2011), in how the cultural attitude developed towards an assessment artefact can determine a teacher's attitude towards it. In Samson and Marongwe, it seems that the teachers were being stricter with marking as they felt that they needed to validate the work which went into the assessment tasks conducted, as well as communicate to students that the assessment tasks were important. It seems that a defensive attitude was developed in this context, as teachers saw the need to defend their own work and give it equal status with the final examination.

Research in other contexts has revealed other factors which may impact the way in which teachers respond to assessment reform. Within the South African context, Lumadi (2011) points out that a similar reform as the one in Malta was introduced in the early 1990s, in which High-Stakes Summative Examinations had their weighting reduced in favour of Teacher-Based Assessment. Lumadi conducted his research with teachers well-over a decade after the reform, so his study may be seen as being more within the context of how assessment systems work, and less within the context of assessment or educational reform. He identifies five factors which affect how teachers respond to this assessment reform, with the strongest variables being the right of information, neuropsychological, feedback, and validity. The variable with the least effect on receptivity was teacher training, however other studies have identified this as being more crucial to assessment reform. (Lumadi, 2011). Lumadi does not specify what he, or his sample, mean by the variable "neuropsychological", so a discussion on it cannot ensue. His study does shed light, however, on how teachers within this context are concerned with the right to information about how the system should work, with this being an essential part of the teachers' receptivity of the reform given that they need to enact the practice pertaining to the policy. Similarly, this study also shows that teachers can be concerned with how valid their

assessment materials are, as well as how they need feedback on the work which they engage in in relation to School-Based Continuous Assessments in order to improve the quality of these assessments. Within this context, these factors gave birth to a negative view of continuous assessment, with this providing a useful insight into factors which can induce a negative receptivity of such reforms.

2.5: The professional development issue.

When assessment reform, or any other reform, takes place, teachers require professional development to deal with the specific demands of it. De Lisle (2016) argues that Continuous School-Based Assessment is a multifaceted and complex educational concept, which combines aspects of summative, formative, and authentic assessment. In turn, different aspects of this may be emphasised by both policymakers, as well as teachers. He argues that the aspects which teachers conceive of as being defining of continuous assessment are what truly determine the types of assessment which continuous assessment will be defined as. The SEC reform leans towards the summative end of the spectrum in how a mark needs to be produced, and in how this affects the over-all mark at the end of secondary education. Teachers' attitudes are what will allow it to contain formative aspects. De Lisle (2016:36) points out, however, that the "Achilles' heel" of continuous assessment is that different teachers may conceive of it in different ways, resulting in different methods of implementation for the same reform. (De Lisle, 2016, p.36). He then goes on to identify the leading factor which results in the failure of continuous, school-based assessments, and this is seen to be undertraining:

"High quality CA training must capture the multi-component nature of CA, with a greater emphasis on formative assessment and use of authentic, performance assessments." (De Lisle, 2016, p.42).

Similarly, Black and William (1998) point out that teachers need more than just assessment resources and instruments, they also need training on how to interpret students' work in a way which is conducive to further learning. The implication of this is that professional development

needs to be administered within the local context considering that this is the first time that parts of SEC will be assessed in schools for subjects such as English.

Assessment reform requires further training for educators. The local reform raises critical areas where such training may be required. Two crucial points that need to be considered are raised by Smith (2011), who argues that there are

“...difficulties in implementing AfL [assessment for learning] as common assessment practice beyond the individual teacher so it reaches the full teaching staff...” (Smith, 2011, p.55).

Within the context of this examination, this matter is more crucial given that the exam is a high-stakes one which is also awarded as a qualification. It would arguably be unfair if teachers engaged in different assessment practices which are for the same qualification; however, there is also an issue with assigning the same task within every classroom in the state as this beats the purpose of classroom assessment, given that one of the purposes of such assessment is to individualize learning for students. This is perhaps a conundrum which policymakers will need to tackle at a later point in the implementation of this reform. A further issue which Smith (2011) raises pertains to the amount of time which it takes to train teachers to implement such a reform. Within the Norwegian context which Smith writes about, teachers underwent approximately 30 months of part-time training to implement a policy aimed at increasing teachers' skill with formative assessment. Within the local context, the author of this dissertation is uncertain as to the amount of training which teachers will be given in order to implement School-Based Assessments for SEC examinations; however, the Directorate for Learning and Assessment is postulating that this training will be going on during the process of data collection for this dissertation. (Smith, 2011, Directorate for Learning and Assessment, 2018). Hopefully, the data collected will lead to an answer for this question within the Maltese reform's context.

The amount of time which is dedicated to teacher training in relation to reform is a crucial aspect of any reform. Central to this is the fact that the policymakers involved in the SEC reform in Malta arguably have different conceptions of time than the teachers who will have to

implement this reform in their daily practice. Hargreaves (1994) calls on the work of the anthropologist Edward Hall in order to argue that there are two broad conceptions of time: “monochronic time” and “polychronic time.” (Hargreaves, 1994, p.102). He argues that policymakers work in contexts which allow them to experience time in a monochronic manner, meaning that they have greater control over their time due to their higher-up positions in the educational system, and hence experience a working day which is highly sequential. He further argues that teachers experience time in a polychronic manner, as multiple things tend to happen within the same measure of time, resulting in less control over their situations. Guskey (2002) points out that one of the aims of professional development is to encourage teachers to change not only their practice, but also their “attitudes, beliefs and perceptions” about an educational concept. (Guskey, 2002, p.382). He provides a model for professional development which is appropriate to mention in relation to the local reform. His model consists of four stages: in the first stage, teachers undergo the relevant training; in the second, they implement the changes brought about through the PD; in the third, they see an improvement in the students’ learning; and in the final stage, they experience a change in their beliefs regarding the practice. The final stage is where the professional development takes place, as this is where one can say that it has been absorbed as part of the educators’ practice. Guskey also highlights the significance of following up any training after PD sessions have been held, as this is where refinement can take place.

The notion of how teachers and policymakers experience time, as well as the time required to conduct effective professional development is not just a theoretical notion. Indeed, research from the 1990s and early 2000s has shown a considerable development in the way in which professional development has been conducted, along with different models being developed. This ties into the teachers’ receptivity of the assessment reform in how teachers, in different contexts, have shown that professional development has been fruitful when it followed an active learning approach, rather than the more traditional workshop type of teaching. (Cohen *et al.*, 2000). This shift is documented in the work of Little (1993) who points out that more conservative models of professional development tended to view the “teacher as troglodyte”, implying that the type of training offered was not in line with constructivist ways of

teaching, and followed a system whereby an appointed expert would provide a talk which dictated what teachers were expected to learn and put into practice. (Little, 1993, p.135). Furthermore, Little (1993) identifies teacher engagement during professional development seminars as being of utmost importance, as well as also identifying the duty of professional development providers to consider the context in which the teachers practise. Naturally, this takes a considerable amount of time to plan and execute. It is worthwhile, however, as research by Harrison (2005) concludes, that in order for teachers' assessment skills to improve, there is the need for "trailing, reflection and thinking", as well as time to discuss, observe, and be observed by peers. (Harrison, 2005, p.262). For teachers to perceive their development as worthwhile, a considerable amount of time needs to be dedicated to it.

Little's research was later corroborated through research by Cohen *et al.* who conducted a study with 1,027 secondary school teachers in the US on their perceived effectiveness of professional development. In their research, they concluded that the best practice for professional development consists of sustained, longitudinal courses, which focus on tackling content through hands-on tasks whilst also being integrated with the teachers' current, or future in the Maltese case, practice. (Cohen *et al.*, 2000). Within the local context, and in relation to English, one would expect the professional development to focus on the actual assessments which teacher will need to do, given the highly practical nature of the reform in place. Similarly, one would also expect the development to focus on specific assessment literacy skills which are required for teachers to be able to perform their new duties with competence and precision

Research by Reid (2007) in the context of writing assessment has yielded intriguing results, which also serve the purpose of backing up the aforementioned argument on the type of professional development which this type of reform entails. In this study, conducted in the Scottish context with teachers from both primary and secondary sectors, Reid (2007) found that the teachers from different cycles of the educational systems conceived of writing assessment rather differently, with this pertaining to how stringent teachers were with applying the criteria used for assessment. The research showed that secondary school teachers were more concerned with matters such as reliability and validity, as well as being stricter with marking.

This leads back to the aforementioned conundrum which bedevils this type of reform: the summative nature of continuous school-based assessment would require the teachers to apply the criterion strictly in the name of reliability; however, doing so would also compromise parts of the formative aspect of this type of assessment in how this would lead to a series of test-type tasks. This aspect will also need to be addressed by any professional development which may occur.

This is of relevance to the research questions of this dissertation in how the teachers will formulate attitudes towards the type of professional development which they will undergo, with this being based on the usefulness of this training in relation to the tasks which they need to carry out. It follows that having data on this topic is relevant as it is related to the teachers' overall receptivity of the reform.

2.6: Self-efficacy and assessment literacy.

This section of the literature review analyses literature which is relevant to the third research question of this dissertation, which aims to examine how efficacious teachers of English are with regard to their assessment literacy skills.

2.6.1: Definitions of self-efficacy and assessment literacy.

Teachers' self-efficacy in terms of their assessment literacy is a crucial component of this reform. Bandura (1997a) argues that:

“Perceived self-efficacy refers to beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments.” (Bandura, 1997a, p.3).

This is relevant to the study in question as, through this reform, teachers are being asked to engage in new behaviours which, traditionally, have not been a part of their job. Bandura further (1997a) argues that:

“People will approach, explore, and try to manage situations within their perceived capabilities, but unless they are externally coerced, they avoid transactions with those aspects of their environment that they perceive exceed their coping abilities.” (Bandura, 1997a, p.14).

This quotation shows why an understanding of the teachers’ perceived current levels of assessment literacy is a crucial component of understanding their overall receptivity of this reform. It follows that if the teachers are efficacious in their ability to handle the changes brought about by the reform, which requires them to do more work in relation to assessment, then their receptivity is more likely to be positive. If they are not efficacious in their assessment literacy, then this could mean that they are perceiving the reform as being coerced and beyond what they can do, leading to a negative receptivity.

At this point, it is crucial to point out what assessment literacy means. Giraldo (2018) argues that, essentially, this “refers to knowledge, skills, and principles in language testing.” (Giraldo, 2018, p.180). Other scholars, however, have maintained that the notion of assessment literacy has implications which go beyond skills pertaining to traditional testing. Jeong (2013) believes that people who are professional language testers have a different understanding of which skills are more important as part of a teacher’s assessment literacy when compared to others who are not professional language testers. In Jeong’s study, language testers tended to believe in the value of test theory, whilst non-language testers considered issues such as classroom assessment strategies and test accommodations to be more crucial aspects of assessment literacy. For the purposes of this dissertation, Giraldo’s broad definition will be used, yet it is worth pointing out that the definition of assessment literacy varies by context.

2.6.2: What does it mean to be “assessment literate”?

Assessment literacy is a crucial set of knowledge and skills which teachers need to possess in order to feel efficacious in their ability to handle the requirements of the reform.

This section of the literature review will elucidate different skills which are often associated with assessment literacy and will focus on ones which are arguably more pertinent to the current situation. A backward design model will be used: the first section will call on literature to explain what someone who is highly assessment literate can do, and will gradually move backwards to define skills which are necessary for someone who is just starting to learn about assessment.

Being assessment literate implies that a person can use several skills related to assessment. Imbar-Lourie (2008) points out that those who have a proficient level of assessment literacy are capable of interpreting experience, as well as conducting research, by calling upon “a body of knowledge and research grounded in theory and epistemological beliefs.” (Imbar-Lourie, 2008, p.396). This means that the person can apply a vast amount of knowledge related to assessment in relation to a task. Furthermore, Scarino (2013) points out that being assessment literate refers to the ability to call upon knowledge pertaining to assessment and be able to work with assessment tasks in a way which is both in line with universal psychometric principles, yet which also considers the context in which the person is working. This person is also then capable of using this knowledge to reflect on their own assessment practice and adapt it when necessary. Such technical skills have often been the domain of the expert language tester; however, with the amount of people directly involved in assessment worldwide, this knowledge is becoming more of a necessity for people who are not experts (Taylor, 2009). This is germane to the reform because teachers are taking on a more active role in their learners’ assessment, and hence, this means that they will be able to implement this reform with greater efficacy if they are knowledgeable of the intricacies of language assessment.

The way in which assessment literacy is explicated in the previously cited work does not delve into specific skills. This section will highlight some skills in the literature which are considered essential to the local reform. One such skill identified by Brookhart (2011) is the ability “to analyze classroom questions, test items, and performance assessment tasks to ascertain the specific knowledge and thinking skills requires to do them.” (Brookhart, 2011, p.8) This skill is important for the reform given that the coursework identified in the syllabus

requires the students to synthesise multiple skills in the same assignment. It is essential for the teachers to be able to ascertain which skills or knowledge are being assessed, so that these inform prior teaching; and also to ensure that students are being assessed on the skills which they are meant to be assessed on. This literature review has also established that one of the advantages of continuous assessment is that teachers can make use of assessment tasks to improve student learning. Brookhart (2011) corroborates this by stating that “teachers should have the skills to provide effective, useful feedback”. (Brookhart, 2011, p.9). This skill is hence essential for teachers to be able to make the most of classroom assessment tasks in order to maximise student learning.

Brookhart’s work provides insights into assessment issues which are relevant to both English language and English literature assessments. Over the years, however, researchers have worked in order to develop a depository of skills which are specifically relevant to language assessment, with these being useful when one considers the particularly high stakes of English Language examinations. One such researcher is Giraldo (2018) who posits sixty-six specific abilities which fall under three broad categories: knowledge, skills, and principles. Amongst these skills, one finds the ability to “design assessments that are reliable, authentic, fair, ethical, practical, and interactive” and also “criticize external tests based on their quality and impact.” (Giraldo, 2018, pp.189-190). These skills are particularly pertinent as they encompass multiple psychometric and sociocultural aspects of assessment which were elaborated on earlier in the dissertation. Being able to efficaciously deal with both paradigms of assessment is seen as a necessary skill for language teachers, as it ensures both scientifically valid and ethical assessment.

2.6.3: Integrating assessment literacy and self-efficacy.

A relationship between literacy and self-efficacy has been reported in the literature, despite there not being a large amount of studies tackling this theme. A scale called “The Self-Efficacy towards Using Alternative Assessment Scale” has been devised by Buldur and Tatar (2011); however, it seems that this scale has not been developed or used considerably in any body of research. A study which sheds some light on the research questions of this dissertation

was conducted by Zhang and Burry-Stock (2003) within the context of a South-Eastern State in the USA. In this study, they point out that “assessment skills” and “assessment practice” are different constructs in how skills are taken to represent what teachers know or are capable of doing empirically, or that which they perceive themselves as being capable of doing, while practice is that which actually takes place in the classroom. (Zhang and Burry-Stock, 2003, p.326). The research being conducted for this dissertation will be limited to “skills”, considering that this reform has not reached the “practice” stage. Zhang and Burry-Stock’s research also sheds light on this study through how they found that measurement training, in other words, training on how to conduct accurate and reliable assessments in class, has a significant effect on how skilled (efficacious) teachers believe that they are when conducting assessments with learners. This type of training was also found to have a higher correlation than years of teaching experience. Similar results were also seen in a study by Mendoza and Arandia (2011) where the researchers found that teachers with training in language assessment had a more positive disposition towards assessment than those who did not. This can partially be related to their self-efficacy in conducting assessments. The implication for this study is that, at least based on this literature, it is probable that teachers with more knowledge and skills in relation to language assessment are more likely to perceive the reform positively, whilst those who see themselves as less capable are likely to experience self-efficacy issues.

2.7: Conclusion.

This literature review has sought to evaluate the proposed reform against literature dealing with similar assessment reforms. The research questions were considered throughout the review in a way which would allow trends to be identified, as well as to identify variables which can inform the data collection tools. The central argument presented is that the teachers are at the heart of this reform, as it is they who will be conducting the Continuous Assessment in whichever form it takes. Given this, it is also they who will be required to take on new responsibilities which are ethical in their nature. The next part of this dissertation will deal with the research methodology adopted.

Chapter 3: Research methodology

3.1: Research methodology and research design.

At the onset, this dissertation was meant to follow a mixed methods methodology, in which a qualitative approach would inform a quantitative end. This was in line with the thinking proposed in Johnson and Onwuegbuzie (2004) who point out that in studies with a mixed methodology approach, the qualitative and quantitative aspects can have both equal and unequal weighting, depending on the researcher's needs. Traditionally, research which has dealt with teachers' receptivity of large-scale reforms (Waugh and Godfrey, 1993; Moroz and Waugh, 2000) and research dealing with teachers' perceptions of assessment (De Lisle, 2015, 2016) have favoured a hard-line quantitative approach. This dissertation did not, and does not, intend on setting a precedent, but opted for a mixed methodology due to the fact that it was the first study to tackle how teachers of English perceive having to assess parts of their students' SEC assessment in Malta. As such, the qualitative aspect, which was meant to be a semi-structured interview, was being conducted to inform a questionnaire which would create generalizable data. This design follows the ideas proposed by Morgan (2007) who points out that "inductive results from a qualitative approach can serve as inputs to the deductive goals of a quantitative approach." (Morgan, 2007, p.71). Whilst the overarching epistemology and ontology of this dissertation were meant to be positivist, "methodological fundamentalism" was rejected in the interest of more valid data. (Waring, 2017, p.19). The original design exploited accepted pragmatism to achieve more positivist ends.

The qualitative aspect consisted of semi-structured interviewing and was intended to then inform a questionnaire. This method is suggested by Gillham (2007) who points out that "careful semi-structured interviewing *prior* [emphasis in original] to the questionnaire being constructed can mean that most of the probable answers are identified." (Gilham, 2007, p.5). The closing down of schools in March led to a number of limitations being imposed on the data collection of this dissertation, specifically when it came to conducting the semi-structured

interviews before the questionnaire. Ravitch (2020) points out that the Covid-19 pandemic has disrupted several research studies in the social sciences, with this being particularly pertinent to qualitative studies which often require face-to-face contact. The pandemic has had an effect on this dissertation, too, as following up requests to conduct research in particular schools became more problematic, given that the researcher could not call these schools to further inquire into the emails which had been sent. The fact that the Maltese educational system experienced a complete transition to online learning has created fresh concerns for teachers and school leadership, meaning that my research may have been (understandably) forgotten about. Considering this, I needed to drop the qualitative aspect of my study and move straight into the quantitative aspect with an online survey. The survey was developed through exhaustive use of the factors and variables identified in the Literature Review, with these then being used to produce “yes/no” questions, multiple-choice questions, and Likert scales were developed to answer the research questions. The questionnaire, as well as its limitations as the sole method of data collection, will be discussed in greater detail in sections 3.3 and 3.6.

3.2: Selection of sites and participants.

The research was conducted with teachers of English practising in the State Sector. The State Sector was chosen for this research for several reasons. The first reason is because this reform is state-initiated, and whilst students wishing to sit for their English SEC examinations in 2023 who attend Church and Independent schools will need to be assessed in the same way, the reform may not have been taking place at the same pace. Different sectors also follow different procedures when it comes to assessment, so the reform may not have been experienced in the same way. The second reason was more practical: the amount of time available for the research was not long enough for three separate sectors to be contacted, for ethical approval to be granted, and for data to be collected and analysed. Bryman (2012) argues that “most of the time decisions about sample size are affected by considerations of time and cost”. (Bryman, 2012, p.197). Considering this, a manageable sample size was chosen.

Initially, six different schools were contacted for the purposes of the interviews of this dissertation. Out of six, one school replied and agreed to participate, whilst the other five did

not reply. Follow-ups via telephoning were unsuccessful, and this issue was further exacerbated when all schools closed in mid-March. Considering this, the interviews needed to be removed from the research design, and the online questionnaire was administered. This was achieved through asking an Education Officer for English to forward the link to the online questionnaire via email. This proved to be advantageous to the researcher, as several issues commonly associated with selecting a representative sample were circumvented and he was able to access the whole population as defined by the research questions.

3.3: Instrumentation.

In order to develop the instruments, which consisted of interview questions and an online survey, the variables which were seen to be relevant in the literature were turned into questions which would lead to data on how teachers are responding to the reform in the local context. Given that the data which was collected was collected solely via the use of the survey, this tool's instrumentation will be discussed here.

The survey was developed using a funnel approach, which is described by Gray (2004) as occurring when the researcher "starts off with a broad set of questions and then progressively narrows down the questions to target specific areas." (Gray, 2004, p.194). In practice, this was applied as the initial research questions were naturally too vague to be presented to the respondents directly. For example, asking teachers how efficacious they are about assessing their students' SEC examinations in class would have been too vague, as both "self-efficacy" and "assessment literacy" are loaded terms which needed breaking down. Using the consulted literature, these were broken down into numerous questions. In the case of the research question referenced here, this was broken down into a Likert scale consisting of various different sub-skills associated with assessment literacy, such as being able to evaluate the validity of an assessment task or being able to give students formative feedback on their work, and these were placed on the survey.

The first two research questions were addressed through closed-ended questions, which were almost all either "yes/no" questions or multiple-choice questions. Gray (2004) points out

that these questions are particularly effective as they provide respondents with structured answers, meaning that they are less time-consuming to answer, and they also allow for comparisons. Concepts which were too complex, such as 'culture', were broken down into indicators. Bryman (2012) defines an indicator as being "something that is devised or already exists and that is employed *as though it were a measure of a concept*." (Bryman, 2012, p.164). For example, culture was broken down into indicators such as "public attitudes towards SEC", "teachers' status", and "teacher accountability". Questions were then posited along a scale of how much of an impact (weak/strong/null) the participants think each one has had on the reform.

Various analyses and reflections were conducted to ensure that the data collection tools were reliable and valid. In order to ensure the face validity of the questions, consultations were made with both an expert (my supervisor), a non-expert (someone with no training in education), and an educational practitioner (a relative who is a teacher, but not of English). Bryman (2012) points out that face validity is intuitive, meaning that this depends on a person's judgment and is not statistical. Whilst Bryman (2012) recommends consulting people with expertise in the field, Gillham (2007) recommends checking with people who are non-experts in the field of study. Gillham's reasoning is based on the premise that a non-expert can provide feedback when it comes to technical jargon, which can arguably be of hindrance to respondents. Had the interviews gone ahead as planned, convergent validity would have also been achievable. Bryman (2012) argues that this is best achieved by having repeated measures "of the same concept through other methods." (Bryman, 2012, p.172). This would have tested the validity of the questions, whilst also allowing for any misgivings to be corrected, however this was not possible given the circumstances. In order to ensure internal reliability in the participants' responses, multiple questions were also posited on the same concept. For example, the issues of reliability and validity in assessment tasks were addressed via both multiple-choice questions, as well as Likert scales at the end of the survey.

3.4: Data collection and recording.

All the data which was collected for this dissertation was done through an online survey, using Google Forms. The online survey was sent to teachers in the third week of April, followed

by a reminder. All items were then coded, with each item given a letter and the possible answers being prescribed representative numbers. The data was then placed on a frequency table, to show the percentages of how the respondents answered the questions. Next, the data was transferred to pie charts and histograms, in order to have a visual representation of the content.

In order to work out the statistical significance of the collected data, Chi-Squared Tests of Independence were conducted. First, the data was placed on a frequency table, then the steps followed were taken from Bryman (2012). Null hypotheses were created, using teachers' genders, years of experience, and qualifications, which were hypothesised in relation to the other variables on which questions were asked in the survey. The p value was then calculated to work out the statistical significance of the findings. Further Chi-Squared Tests were conducted between variables that are often tackled in pairs in the literature; for example, teaching and learning, and reliability and validity. Cramer's V was then calculated when a significant relationship was found to see the strength of the relationship between the two variables. The implications of these findings were then discussed in relation to literature on the topic.

3.5: Methodological assumptions.

The dissertation assumed that the participants were honest and answered the questions to the best of their knowledge. Gillham (2007) points out that there is a tendency for "people not to take questionnaires seriously" and that this can lead to issues in the reliability of the data collected. (Gillham, 2007, p.13) This dissertation is assuming, however, that the participants answered the questions to the best of their knowledge.

This dissertation also assumed that the participants were considerably knowledgeable of the reform taking place, given that it will have an impact on their working lives. As per Gilham's (2007) recommendations, however, the cover letter of the questionnaire clearly specified which reform was being discussed, as well as the rationale behind the dissertation.

3.6: Limitations.

This section will discuss two topics: the specific limitations brought about due to the Coronavirus Pandemic, as well as the general limitations associated with using Online Surveys to collect data

3.6.1: Limitations of removing the qualitative aspect from the dissertation.

In section 3.1, it was established that this study was meant to follow a mixed methods design; however, due to time limitations brought about by the Covid-19 pandemic, the qualitative aspect needed to be removed. Gillham (2007) points out that prior to conducting a questionnaire, it is always beneficial to conduct a handful of semi-structured interviews with members of the target sample in order to clarify any misgivings which the researcher/s may not have noticed during the planning stage of the research. Furthermore, Bryman's (2007) research attests to this in how a number of prolific academics, who were interviewed on their perceptions of the challenges of mixed methods research, argued that not integrating quantitative and qualitative methods can often lead to the value of a piece of research being decreased. McKim (2017) also points out that the qualitative aspect of a mixed methods study increases the validity of the data collected, given that participants are freer to express their opinions and are not confined by the options provided by the researcher. These arguments lead the researcher of this study to conclude that dropping this aspect of his dissertation has resulted in a loss of validity, which was undesirable, yet inevitable due to time constraints and the additional difficulty in recruiting participants created by the pandemic.

3.6.2: Limitations of the online survey as a data collection tool.

Whilst the online survey is a convenient data collection tool which allows for great quantities of data to be collected in a short period of time, it also presents itself with several disadvantages which are unique to it when compared to traditional paper-based questionnaires. This section will discuss selected issues with online surveys, and how these issues were overcome in the research.

One issue which is pertinent to many surveys is that response rates are generally low. In their study, Deutskens *et al* (2004) found that the two best ways to increase the response rates of their survey were to send reminders and include incentives. Whilst this dissertation did not include any incentives for participation, not least because this is not deemed ethical, a reminder was sent out two weeks after the initial moment of communication. Their study shows that whilst early follow-ups, sent within one week, have a greater influence on the amount of responses received when compared to sending reminders two weeks later, this difference is not statistically significant (Deutskens *et al*, 2004). The length of the survey was seen to have a bearing on the amount of responses received, with surveys consisting of 20 items or less having a higher response rate (Deutskens *et al*, 2004). This may have proven to be of hindrance to this dissertation's data collection, as my survey consisted of more than thirty items. Whilst the response rate for online surveys is generally low, Hunter (2012) points out that the "subject matter, clarity of the covering letter, and the ease and convenience of the chosen response method are examples of other factors that will play a part" when it comes to the amount of responses gathered. (Hunter, 2012, p.15). The researcher believes that the subject matter is of interest, given that it is a contemporary topic in Maltese education, and that the method by which the respondents were asked to reply is convenient, given that all that was asked was for them to click on hotspots.

A further issue which is identified in the literature is that one often does not know who is filling in the online survey. Van Selm and Jankowski (2006) point out that very often online surveys are left in the public domain on the Internet, and that this leads to sampling errors. The researcher of this dissertation did not distribute the link to the survey via any means other than the email sent to the Education Officer in order to forward the link to the participants. With this being said, it was impossible for the researcher to know where this link ended up after it was forwarded to the participants. The issue of not knowing who filled in the questionnaires could have confounded the data in ways which the researcher could not circumvent.

3.7: Establishing credibility.

For the credibility of the study to be established, several tests and reflections were conducted throughout the data collection phase. The survey questions were piloted with

several people with different levels of expertise in relation to the field to ensure the face validity of the questions.

3.8: Ethical considerations.

The study and instruments used were deemed ethical by the Faculty of Education's Research Ethics Committee (FREC). This means that there was no possibility of harm coming upon the participants of this study. All data were collected in line with the GDPR's guidelines, meaning that IP addresses were not trackable by the researcher and the participants enjoyed full anonymity throughout. All data will be destroyed within three months of the dissertation's completion. Also, the participants were instructed not to impart any personal information or information regarding their place of work to ensure anonymity. All participants complied and no information regarding who the participants are or where they work was gathered.

3.9: Restatement of the purpose.

This chapter has sought to highlight the research methodology adopted to examine how teachers are responding to now having to assess parts of their students' SEC examination, with this being the major thrust of the study. The first secondary research questions were formulated as follows:

RQ2: Which factors/ variables have a strong relationship with English teachers' receptivity of the local assessment reform?

RQ3: How do teachers perceive that this change will affect their working lives, as teachers of English in the Maltese educational system?

These research questions were explored via the use of closed-ended questions in the online survey, with these questions comprising of "yes/no" questions and multiple-choice questions. The final research secondary question dealt with teachers' self-efficacy regarding their assessment literacy, and was as follows:

RQ4: How efficacious do teachers of English believe that they are when it comes to assessing the language skills and knowledge of their learners for the purposes of a high-stakes examination?

This research question was investigated by asking the participants to rate their self-efficacy on a Likert scale ranging from one to four on several skills which pertain to assessment literacy. The following chapter will present and discuss the findings from the data collected.

Chapter 4: Findings and Discussion.

4.1: Introduction.

In this chapter, the results of the online survey are presented and discussed. The purpose of this chapter is to evaluate the findings against that which was established in the Literature Review. The first two sections of this chapter deal with who the participants are, their gender, experience, and qualifications, whilst the rest of the chapter discusses the concepts and variables identified in the Literature Review. Furthermore, Chi-Squared Tests of Independence were conducted for each variable using the teachers' genders, years of experience, and qualifications. A significance threshold of 0.05 was used to determine dependency, following Bryman's (2011) arguments. Further to this, when significance was established, Cramer's *V* was calculated. Following Bryman's arguments, a *p-value* generated through Cramer's *V* of less than 0.25 is considered as weak, whilst one above 0.25 up until 0.50 is considered as moderate. Relationships above 0.50 were considered strong. Only tests which provided statistically significant data are reported in this chapter.

4.2: The sample and response rate.

The survey was sent to all 250 teachers of English currently working in the state sector, of whom 212 (85%) are female, and 38 (15 %) are male (Figure 1). Eighty-one teachers answered the online survey registering a response rate of 32.4%, of which 83% were female and 17% were male. Proportionately, more males than females participated in the survey. The following pie chart visualizes the ratio of male to female teachers:

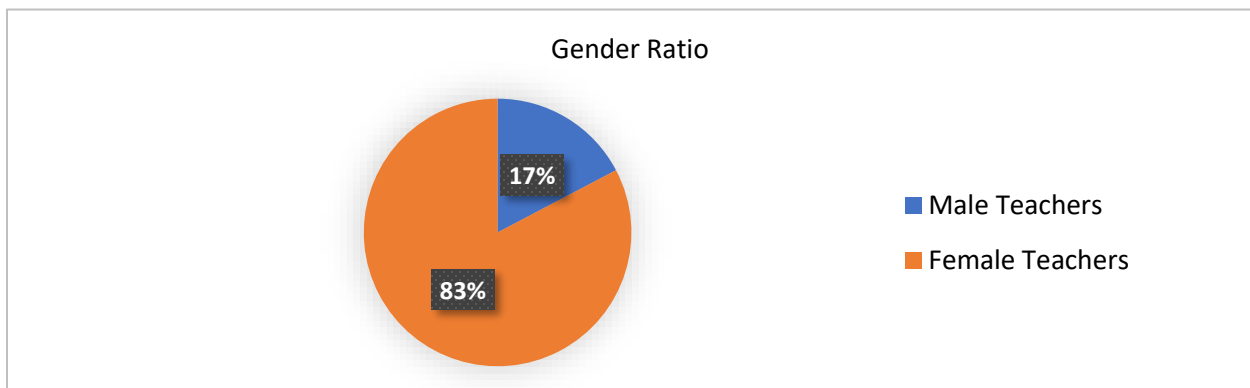


Figure 1: Participants' genders.

4.3: Participants' years of experience and qualifications.

The profile of the respondents in terms of years of experience is shown in the bar graph below. (Figure 2).

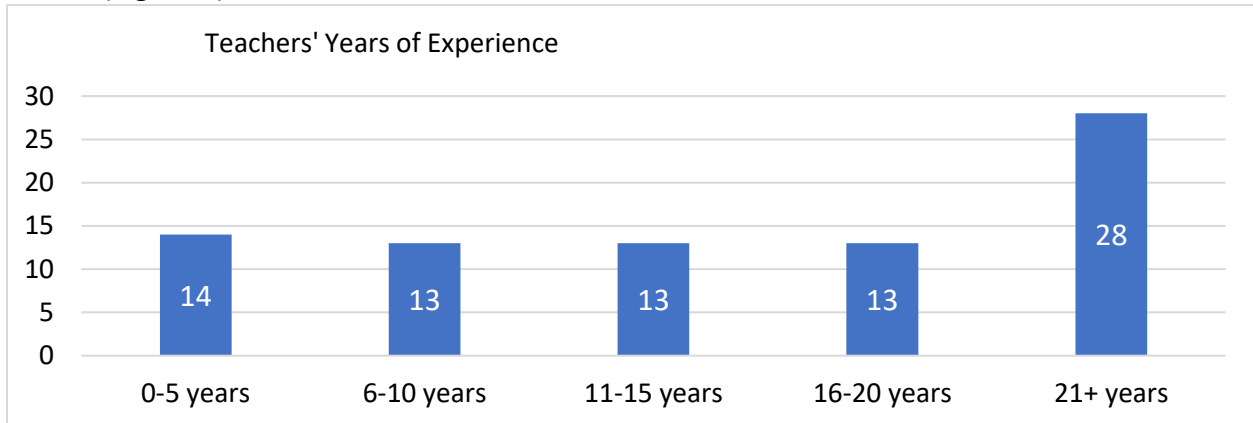


Figure 2: Participants' years of teaching experience.

This data shows that, within the sample, 35% ($n=28$) of the teachers had 21 or more years of experience when it comes to teaching English. The bands which consisted of 6-10 years, 11-15 years, and 16-20 years all represented 16% ($n=13$) of the sample population each. The band which consisted of teachers with 0-5 years of experience represented 17% ($n=14$) of the population. This is indicative of a sample population which is considerably experienced when it comes to teaching English. The average teacher in this study has approximately 11-15 years ($M= 3.38$) of teaching experience, with a Standard Deviation (SD) of 1.5 being reported. Given that the category in the middle of the bar chart is where the median falls and there is a Standard Deviation of 1.5, each group can influence the results gathered on tests of statistical significance.

When it comes to qualifications, a diverse amount of responses was gathered. The data shows that the most common qualifications are the "Bachelor's Degree in Education (B.Ed.)" ($n=25$) and the "Bachelor's Degree not in Education + P.G.C.E. ($n= 24$)." (Figure 3). These two sets of qualifications make for 31% and 30% of the total sample population, respectively.

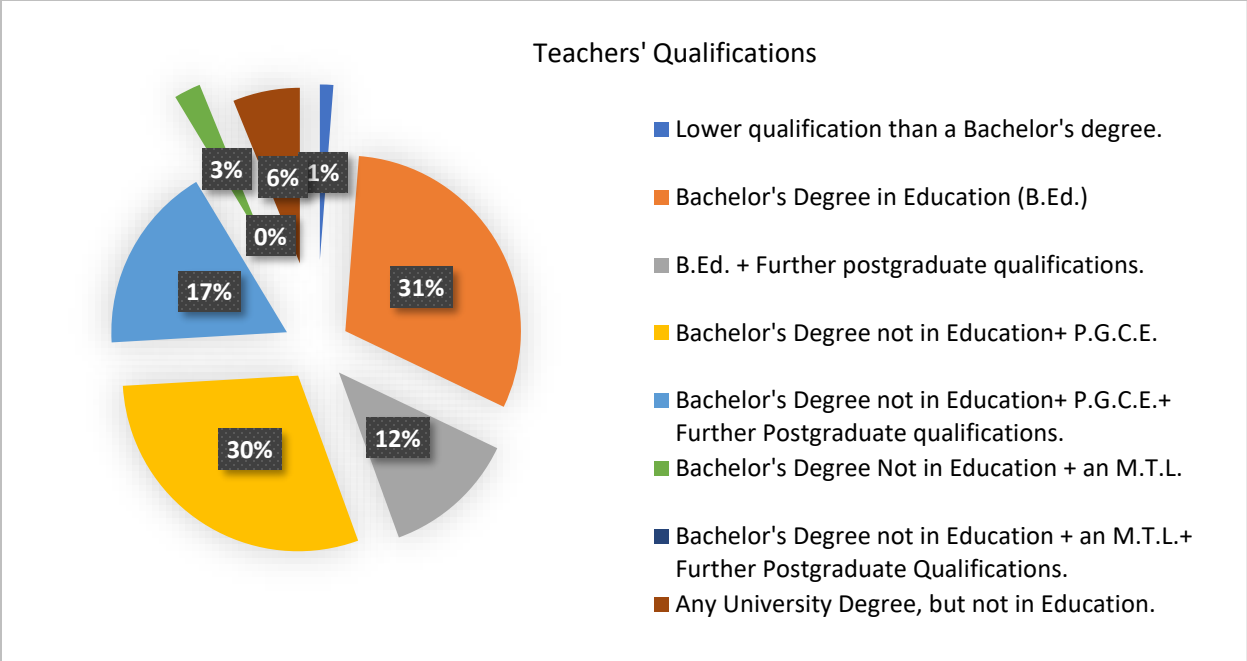


Figure 3: Teachers' qualifications.

4.4: Teachers' receptivity of the amount of Information provided by the authorities.

This question sought to examine how the teachers of English felt about the amount of information provided to them by the authorities. The purpose of this question was to see whether the teachers are perceiving the information given to them as being sufficient for them to implement the reform. The data shows that only 12% ($n=10$) of the participants think that they have all the information which they need, whilst over half (53%, $n=43$) think that they have enough but would still like more. Finally, 35% ($n=28$) think that they do not have enough information with regards to the policymakers' proposed reform. This reflects a need for further action to ensure that all students have a teacher who is well-informed and confident in their knowledge of this reform. In the Literature Review, Brown's (2004) study was utilized in order to show that teachers may not always understand the intended implications of a reform and Lumadi's (2011) research was used in order to posit the argument that when dealing with such complex assessment reforms, teachers need to have their right for information met. In light of the local reform and the data gathered, it follows that further clarification is required for the teachers to perceive themselves as having the sufficient amount of information needed to complete the duties related to this reform. Given that no significant relationship was found

between gender, years of experience, and qualifications, it follows that the issue is random and that it is not one particular group of teachers who feel that they need more information.

4.5: Level of involvement in the reform.

This item sought to collect data regarding how involved the teachers felt in this reform. (Fig. 4). Fullan (2003) points out that top-down reform is not as effective as reform which originates at the bottom of the educational ladder, and this led the researcher to conclude that it is appropriate to ask teachers how involved they felt they were in this reform.

This data shows that only a minority of 5% ($n=4$) reported that they were highly involved in the reform. 57% ($n=46$) were slightly involved in this reform, whilst 38% ($n=31$) were not involved at all. The mean value (m) for this item was found to be 2.33, which shows a tendency of slight involvement and non-involvement. (Figure 4).

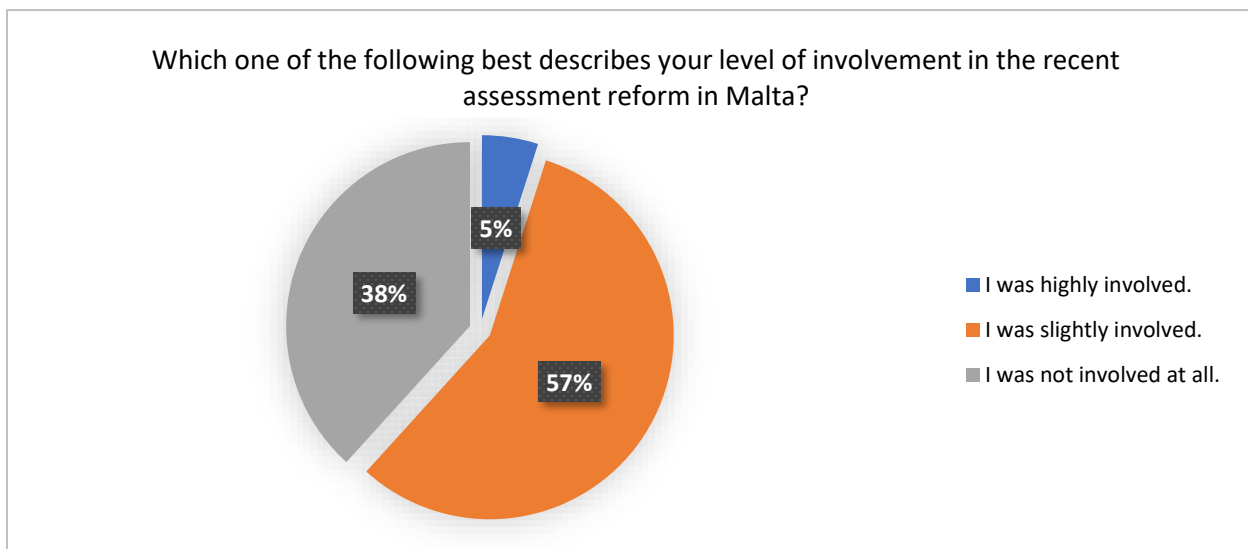


Figure 4: Teachers' perceived level of involvement in the reform.

Whilst the majority of the participants had some form of involvement, a considerable minority were still not involved in the process of reform, implying that non-participation is clearly present. Through Marion's (1999, as cited in Fullan, 2003) theory of auto-catalysis, it is arguably the case that a considerable proportion of the participants in this study are not actively involved in bringing about the change in this reform; but rather, it is the changes which are happening elsewhere causing changes in the participants' behaviour. Given that no significant relationships were found between the individual participants' profiles and their responses for

this item, it follows that this is random and an individual group within the sample cannot be influencing the gathered results. Following Fullan's (2003) argument that bottom-up reform is more effective than top-down reform, it is sensible to argue that further involvement from teachers will serve to further flesh out the reform and help it to achieve its aims.

4.6: The role of public perception in the reform.

This question intends to examine how teachers perceive public perception to have influenced this reform. The Literature Review established that public perception often plays a role in assessment reform, and hence the researcher thought that it was sensible to ask the participants about this concept. An overwhelming majority of 83% ($n=67$) think that public perception influenced this reform, in varying degrees. (Figure 5).

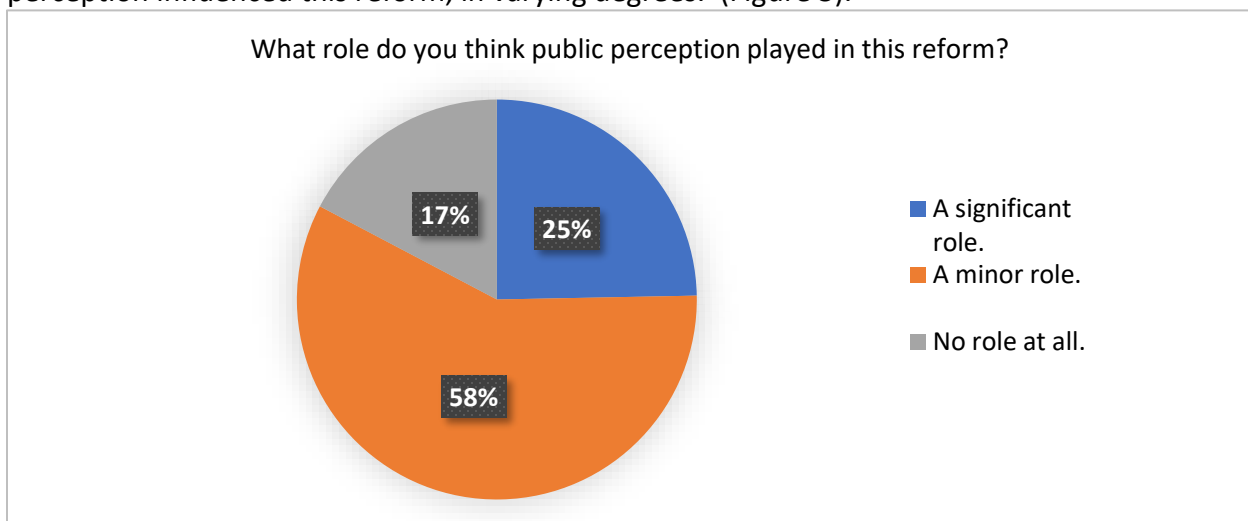


Figure 5: Teachers' perceptions of the role of public perception in this reform.

Considering Gladwell's Broken Windows Theory (as cited in Fullan, 2003), this data is positive in how it shows that the majority of the teachers perceive the general public as caring about how Maltese youth are assessed. It is worth pointing out, however, that this is the participants' perspective on this issue and may not hold true if data were to be collected from the general public itself. Whilst Mee (1998) argues that the general public in the Singaporean context were critical of Teacher-Based Assessments, it seems that the participants of this study see the general public as being one of the catalysts of the reform as the majority identify the public as influencing this reform. Further insights, such as those in Davison's (2004) study, on how the general public perceives teachers' work will only be possible in the local context after this

reform is implemented and being enacted, given that the teachers have not yet been required to assess students for their SEC. This item does, however, show that the participants see the general public as having an influence on the reform in question. Statistical tests found no relationships between the participants' profiles and responses for this item, indicating randomness.

4.7: The influence of high-stakes international examinations.

This item dealt with how teachers perceive Malta's below average scores on international examinations (such as PISA and PIRLS) as influencing the reform. The data shows that the majority (96%, $n=78$) of respondents think that Malta's below average scores on international tests had an influence on this reform. Only 4% ($n= 3$) perceive this factor as having had no influence on the reform (Figure 6).

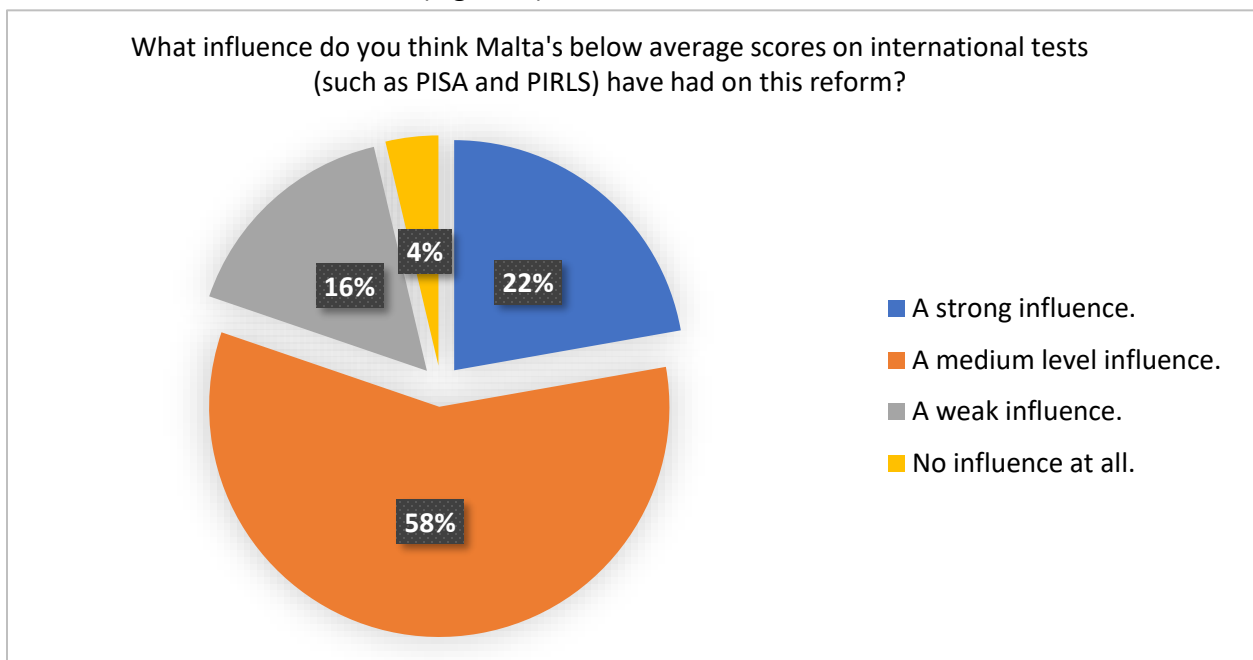


Figure 6: Teachers' perception of the impact of the high-stakes examinations on the reform.

The fact that the majority of the participants think that these examinations had an impact on the reform in question corroborates the argument made in the Literature Review, that the low performance is a cause of concern and that these examinations are in fact perceived by teachers to have an impact on education. No significant relationships between gender, years of experience, and qualifications were found, and this reveals that this is a general sentiment in the surveyed population. Whilst Goldstein (2004) and Mifsud (2016, for MEDE) point out that

such international examinations are riddled with issues of linguistic biases, the point stands that, at least in this population, the examinations are seen to be impactful.

4.8: Teachers' workload.

Two questions on the survey sought to examine whether the teachers perceived their workloads as increasing because of this reform. The overwhelming majority of respondents (96%, $n=78$) perceive this reform as increasing their workload. 4% ($n=3$) think that their workload will remain the same, and none of the participants chose the option which would represent a decrease in their workload and, considering this, no Chi-Squared Tests were conducted. This conforms with literature on the topic, as various researchers have found that Continuous Assessment programmes which require assessment to be done in class do increase teachers' workload when compared to educational systems which do not adopt such an assessment system. For example, Nsibande and Modiba (2012) found that teachers saw Continuous Assessment as consuming more time both in terms of preparation, as well as in terms of time taken up in class to conduct the assessments. De Lisle (2016) also points out that Continuous Assessment reforms often necessitate a "shared redefinition of the teachers' role or [the assessment system] may lead to an intensification of teachers' work." (DeLisle, 2016, p.42). This survey shows that, at least in the group surveyed, an increase in the workload is being perceived as a reality by most of the participants.

The participants were also asked whether, as teachers of English, they see their workload as being greater than that of teachers of other Secondary School Subjects. 95% ($n=77$) think that their workload will not be the same as that of teachers of other subjects, whilst the rest think that their workload will be the same. The final respondent has been listed as "Other" as his/her response was given under the "Other" option, and this participant did not provide an answer which falls under "Yes" or "No". This participant's answer stated that it might be the same as that of other (non-English) language teachers. Under the "Other" section provided on the item, some participants ($n=6$) explained their reasoning, with some citing the fact that English teachers in Malta need to prepare students for two SEC examinations which, in the context of this reform, means that they need to conduct assessments for two separate subjects with the same class of learners. Further to this, teachers also cited having heavier lesson loads

than teachers of other subjects as part of their reason for choosing the “No” option. From these two items on the survey, it is evident that the surveyed English teachers perceive their workload as increasing and that their workload is not the same as teachers of other subjects. This implies that the whilst the reform may lead to a greater workload for all teachers, English teachers are disproportionately affected.

4.9: Teaching and learning.

The questions which followed dealt with the inter-connected topics of teaching and learning. In the Literature Review, it was argued that this reform fits into an educational system which aims to increase students’ learning, with this being noticeable in the goals listed in the *Framework for the Education Strategy for Malta 2014-2024*. (MEDE, 2014). This policy lists reducing achievement gaps and “raising the bar in literacy...and increasing student achievement” as its main priority, so it is sensible to argue that the series of reforms in place after the publication of this document aim to achieve this goal. In this part of the survey, the participants were asked about how they think this reform will affect their teaching methods, as well as their students’ learning. The item on teaching methods will be presented first, followed by that of students’ learning.

Over half the teachers (54% $n=44$) think that this reform will require significant changes to their current teaching methods, whilst 43% ($n=35$) think that it will require minor changes to their teaching methods. Only 3% ($n=2$) think that the reform will require them to make no changes to their teaching methods. (Fig.7).

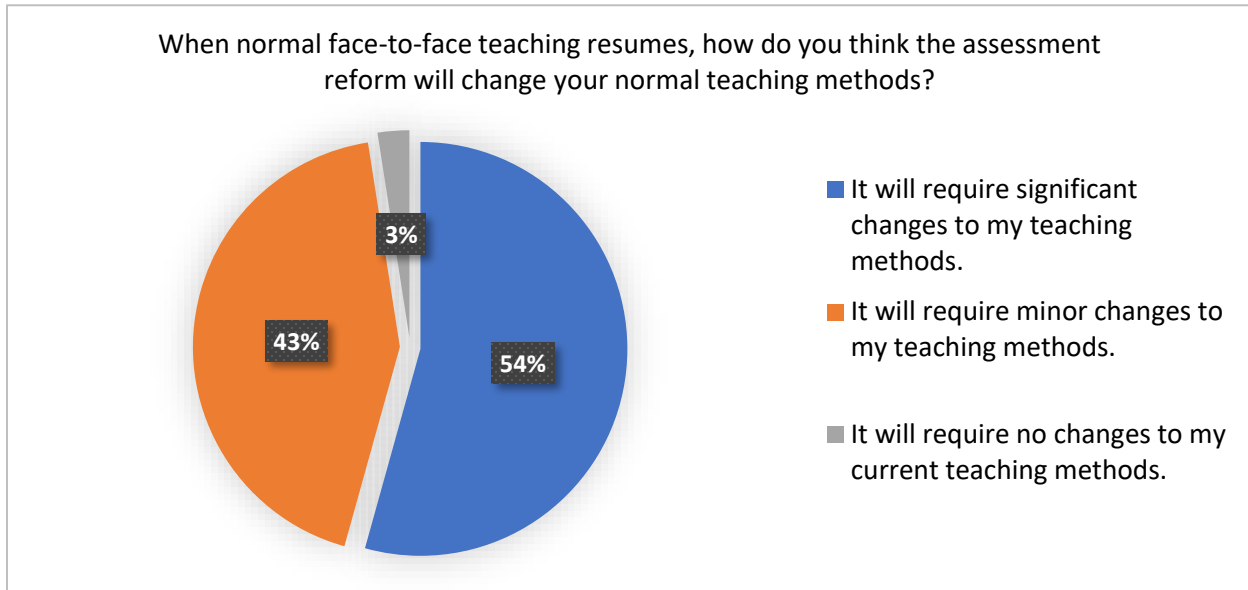


Figure 7: Teachers' perception of how this reform will affect their teaching methods.

No statistically significant relationship was found between responses for this item and teachers' profile information. Given the responses provided for the item which dealt with International High-Stakes Examinations, this is not necessarily a negative outcome. Black and William (2010) argue that in order for student learning to be improved, changes need to happen in the classroom before they happen elsewhere, so this anticipated need for change may result in better learning if the reform leads to better teaching methods.

The previous item established that the majority of the participants perceive this reform as requiring them to change their teaching methods. It would be presumptuous, however, to argue that a change in teaching methods will lead to an increase in student learning. The next question on this survey asked the participants to consider how they think the reform will impact their students' learning. The most common answer (46%, $n=38$) for this question was that the reform in assessment will not increase students' learning, while 38% ($n=31$) believe it

will lead to a moderate increase, which shows a rather sharp division in the participants' responses for this item. (Fig. 8).

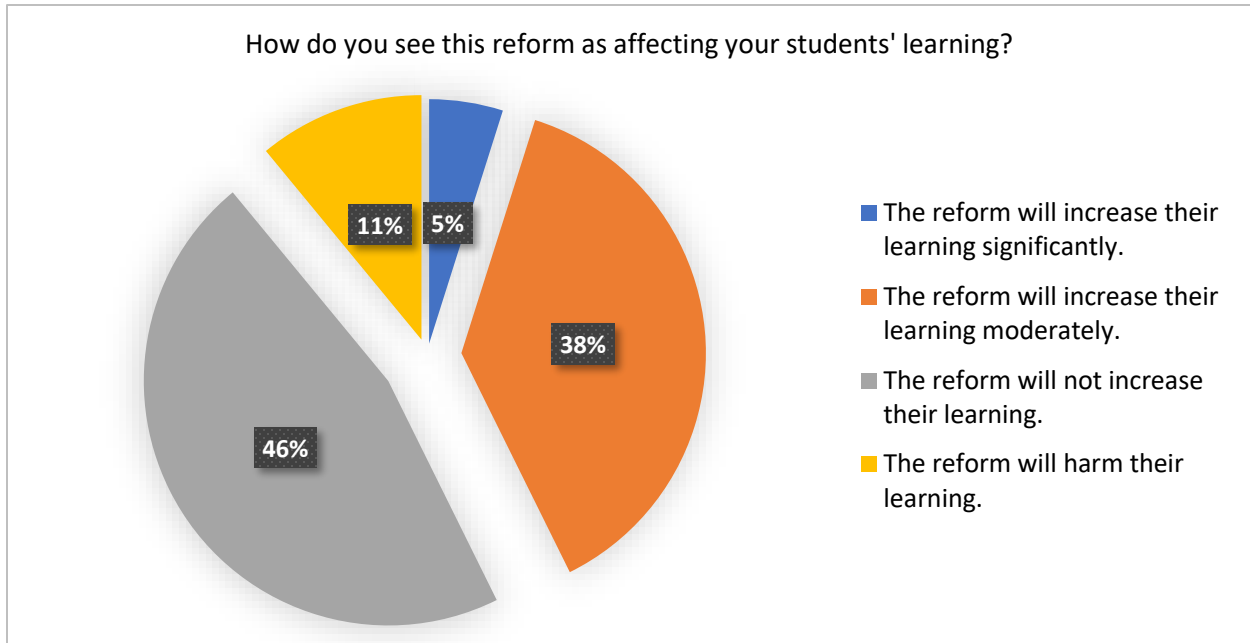


Figure 8: Teachers' perception of how this reform will impact students' learning.

A Chi-Squared Test was conducted using gender and the responses for this item, and significance was established ($p=0.004$). Cramer's V was calculated and a correlation coefficient of 0.4 was found, indicating a modest relationship between the two variables. Whilst the male teachers of English in this study are more likely to see the reform as harming students' learning, the female teachers are more likely to select moderate responses. Another Chi-Squared Test found no dependency between teachers' experience and how they perceive this reform as affecting their students' learning ($p=0.07$). A final test was conducted using the teachers' qualifications and their response for this item, and the result was significant ($p=0.02$). Cramer's V was also calculated, and the correlation coefficient was found to be 0.4, indicating a moderate relationship as well. The data showed that more highly qualified teachers tended to view the reform as having a greater, positive, impact on learning whilst the less qualified teachers were more pessimistic about the reform's influence on learning.

Another test was conducted to see if there is a relationship between teachers' views on how this reform will change their teaching methods and their views on how they see this

reform as impacting students' learning. The *p-value* was found to be 0.07, which implies that there is no significant relationship between teachers' views on how this reform will affect their teaching methods, and how teachers see this reform as affecting their students' learning. Given the discussion in the literature review on how this reform is arguably intended to increase students' learning through better quality teaching, the lack of significance is rather difficult to contend with. This is revelatory of the surveyed population seeing the reform as requiring a change in teaching methods, without it having an effect on students' learning.

4.10: Issues of reliability, validity, and credibility.

The literature review of this dissertation discusses how crucial reliability and validity are in ensuring fairness in any assessment. Given that the scope of this research is to understand how teachers are responding to having to engage with new assessment events in their classrooms, questions were asked regarding how they feel about the reliability and validity of assessing parts of students' SEC examination through Teacher-Based Assessment. Black (2001) points out that reliability is achieved when the student achieves the same result for the same assessment task in different contexts.

Considering the literature on this topic, the results are somewhat concerning. Wyatt *et al.* (2010) point out that Teacher-Based assessment often faces issues when it comes to reliability; however, the data gathered show a considerably low level of confidence in both one's own ability to assess students' learning reliably, as well as in that of other teachers. Just under two thirds of the teachers (63%,*n*=51) are moderately confident in their own ability to

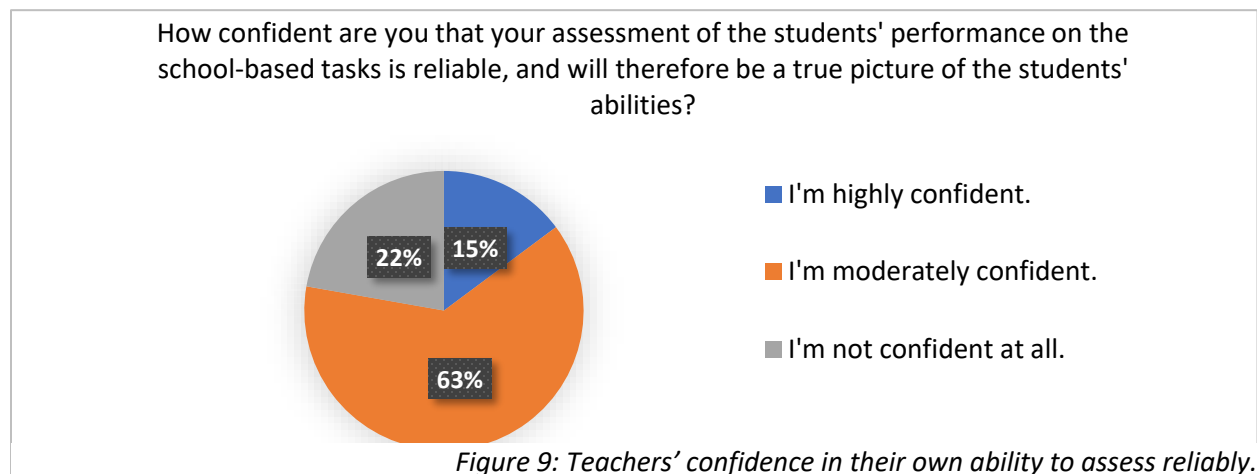


Figure 9: Teachers' confidence in their own ability to assess reliably.

assess students' work reliably, whilst 15% ($n=12$) are highly confident in their own ability to assess reliably, and 22% ($n=18$) are not confident at all (Figure 9).

Two thirds (66%, $n=53$) are moderately confident in other teachers' ability to assess students' work reliably, but only 7% ($n=6$) are highly confident in other teachers' ability. A staggering 27% ($n=22$) are not confident at all in other teachers' ability (Figure 10).

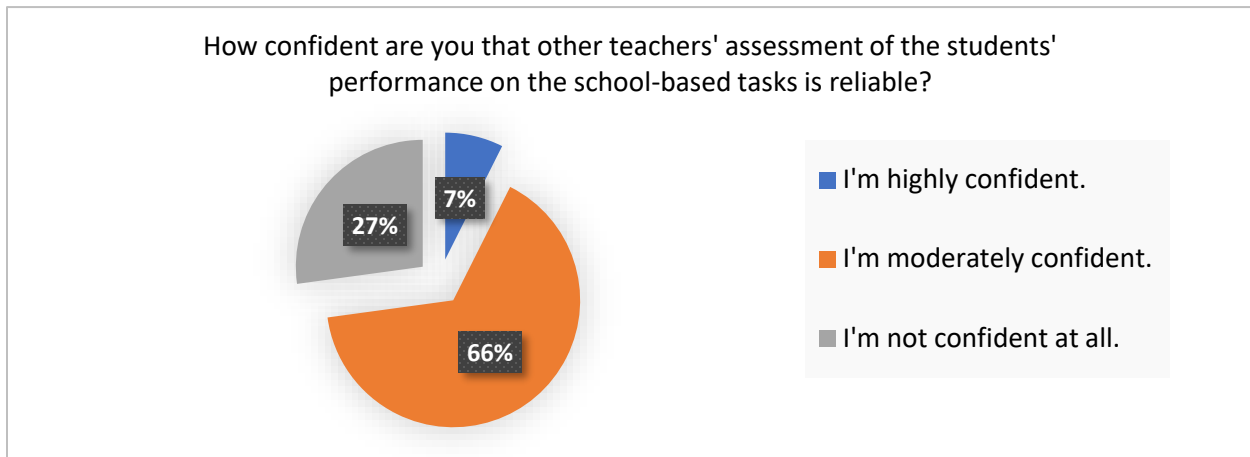


Figure 10: Teachers' confidence in other teachers' ability to assess reliably.

Literature on the topic of reliability reveals that Teacher-Based Assessment is often riddled with problems when it comes to reliability. For example, Wyatt *et al.* (2010) point out that whilst validity is generally secured in Teacher-Based Assessment, reliability is often compromised. The data collected is in line with this research, as it shows that the participants are not highly confident in this method of assessing learning. Chi-Squared Tests for these items found no significant relationship between the responses for this item and the teachers' genders, experience, and qualifications. Another Chi-Squared Test was conducted using the responses to these two items on the survey, and a significant relationship was found ($p=0.0005$). Cramer's V was calculated using these two variables, and a correlation coefficient of 0.35 was found. This shows that the way in which the participants view their own ability to conduct assessments reliably, and the way in which they view their counterparts' ability to assess reliably, are moderately related to each other. Considering this, it is sensible to argue that action needs to be taken to increase the participants' confidence in both their ability to

assess reliably, as well as have faith in other teachers' assessments, considering that the qualification in question is a national one.

Participants were also asked how valid they think the proposed Teacher-Based Assessments are when it comes to assessing students' learning. 11% ($n=9$) of the participants are highly confident whilst nearly a quarter of the teachers (24% ($n=19$)) are not confident at all in this means of assessment. Almost two thirds (65% ($n=53$)) are moderately confident. (Figure 11). The mean (M) answer for this question was 2.12, with a Standard Deviation (SD) of 0.6, showing that most of the participants "sat on the wall" by choosing the response in the middle.

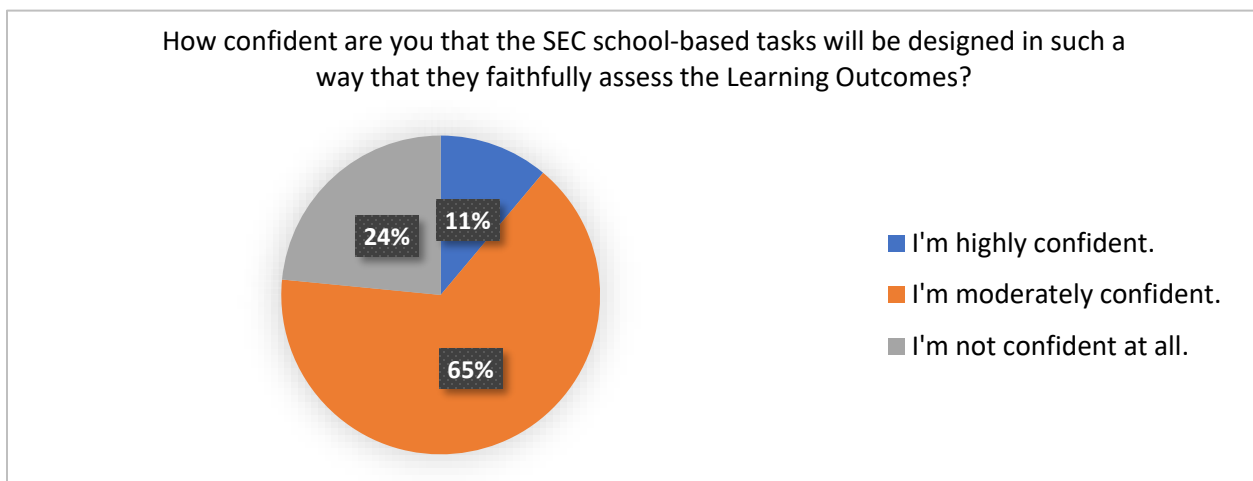


Figure 11: Teachers' confidence in Teacher-Based Assessment as a valid way of assessing students' learning.

Harlen (2005b) points out that validity and reliability are closely related in Teacher-Based Assessment: validity generally increases whilst reliability decreases; however, the data gathered here show that the responses are roughly on the same level, with just under or just over two thirds feeling moderately confident, and the rest picking an extreme option. The low level of responses in the option which would indicate high confidence is a cause for consternation, given that invalid assessments will not produce results which show whether or not Maltese learners are receiving their core entitlement. No significant relationships were found between gender, experience, and qualifications for this item, showing that the issue is spread across various sub-groups amongst the participants. Ideally, all participants would be highly confident in both the validity and reliability of the tasks undergone by their students for the purposes of the SEC.

Another item dealt with how the participants perceive this reform as affecting the SEC qualification's credibility. Over half of the participants (58%, $n=47$) think that the SEC qualification will lose some of its credibility, whilst 37% ($n=30$) think that this will remain the same (Figure 12).

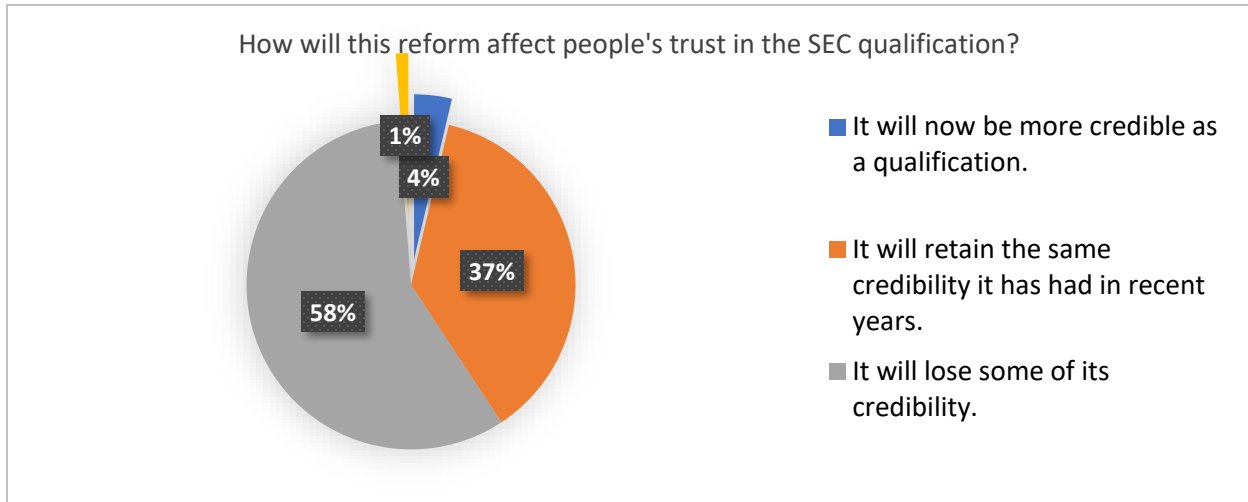


Figure 12: Teachers' perception on how the reform will affect the SEC qualification's credibility.

The majority of the sample think that this reform will have a negative effect on the credibility of SEC qualifications, which is concerning given the arguments made in the Literature Review on how the SEC qualification is a gateway to employment and higher education. Isaacs (2014) points out that when an examination with such high stakes loses parts of its credibility, there are negative repercussions on both the examinees and society at large. No statistically significant relationship was found for this item and teachers' gender, experience, and qualification. Nor was a significant relationship found between this item and the item for validity, showing that the two items are not seen as being related in this context. This being a general sentiment makes the findings even more concerning given that they are not located in a specific group. This issue would have benefited from further qualitative research, which would explore why the participants think that this qualification will be losing some or all its credibility, yet the fact the majority chose that it will lose its credibility is disconcerting.

The final question for this section deals with who the teachers think should prepare the actual tasks to be used for students' assessment at SEC level. In the Literature Review, Wyatt *et al.*'s argument was presented to show that even if central authorities provided the tasks for the teachers, issues of reliability would still be present given that teachers may have different conceptions of what is considered satisfactory performance, and also because of biases. Just over half the teachers (52%, $n=42$) think that the tasks should be provided by a central authority, whilst 42% ($n=34$) think that the tasks should be prepared by teachers working together in the same school (Figure 13).

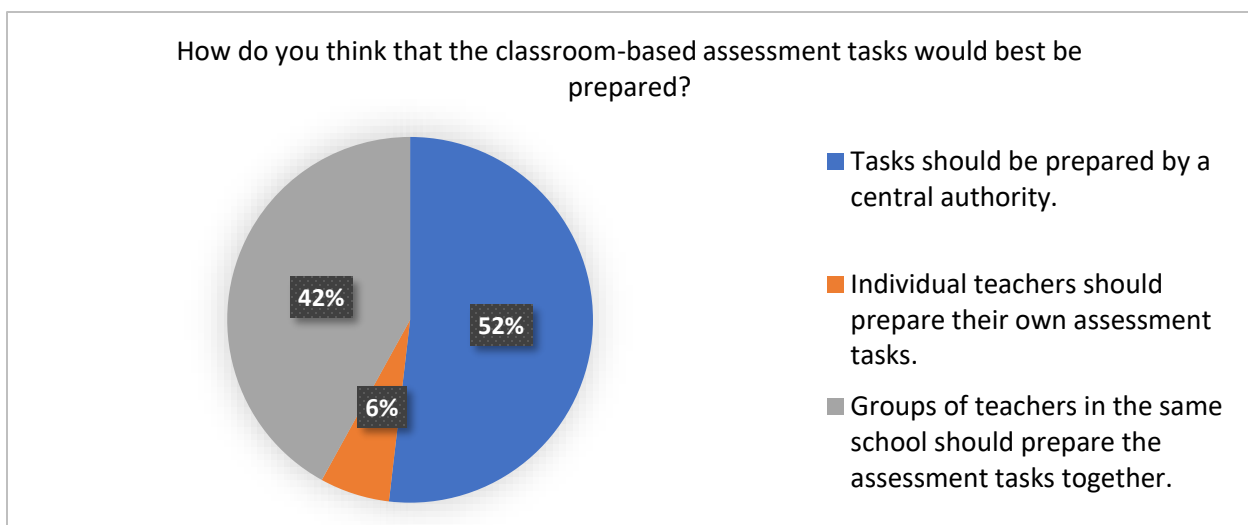


Figure 13: How teachers think the assessment tasks should be prepared.

One plausible explanation for this result could be that the teachers perceive the central authority as being more capable of providing valid assessment tasks, although issues in their reliability would remain pertinent. Lumadi (2011), for example, shows that attitudes towards authority play a key role in how teachers view Continuous Assessment and its relevant procedures. An explanation for how 42% think that the best option is for groups of teachers to prepare the tasks together is seen in how the teachers may feel more confident in the reliability of the tasks, due to inter-rater reliability in both the drafting and correcting process, if this is done in groups rather than individually. Consequently, however, this reliability may compromise the validity of the task if students in different tracks or on different learning programmes are given the same task. This links back to the multiple conundrums discussed in

the Literature Review in how the tasks cannot be both provided by a central authority and prepared by the teachers in a school, and is hence an issue which the relevant stakeholders may need to grapple with.

4.11: Teachers' perceived roles, rapport, and accountability.

This section deals with how the surveyed sample feels that the reform will affect their daily working lives as teachers, as well as their position as teachers within society in general. The literature review established that there is a relationship between assessment reform and a teacher's position within a given culture, and this part of the survey attempted to understand how teachers are perceiving this reform as changing their professional lives.

The first question sought to establish whether the participants saw their roles as teachers changing. The data shows that a slim majority of 54% ($n=44$) think that this reform will change their roles as teachers, whilst 46% ($n=37$) think that their role will remain the same. This sharp divide shows that over half of the surveyed population thinks that what it means to be a teacher is changing because of this assessment reform. The responses for this question are reflective of literature which shows that including course work in High-Stakes Assessment creates tension in roles as, whilst half of the participants think that their role is changing, the other half do not. In the context of science education, Gioka (2009) found that teachers either took on the role of examiners or meddled between being teachers and examiners. Her research contains no data indicating that teachers who work under such assessment schemes remain solely teachers. It would be interesting if this question were to be posited to teachers in future research after the reform has been implemented. These results are also in line with van den Berg's (2002) argument, that educational reforms bring about role ambiguity, with this being seen in the sharp divide in the responses for this item. As a whole group, the answer is quite ambiguous as it is split down the middle, leaving the researcher unable to provide an answer as to whether what it means to be a teacher will change because of this reform. Chi-Squared Tests also found no significant relationships between teachers' genders, experience, and qualifications and their responses for this item.

The next question on the survey asked the participants how they think this reform will affect their rapport with their students. The absolute majority of respondents (80% ($n=65$)) do not see this reform as having an effect on their rapport with their students. 11% ($n=9$) think that it will have a negative effect and 9% ($n=7$) think that it will lead to improvements in the relationship (Figure 14).

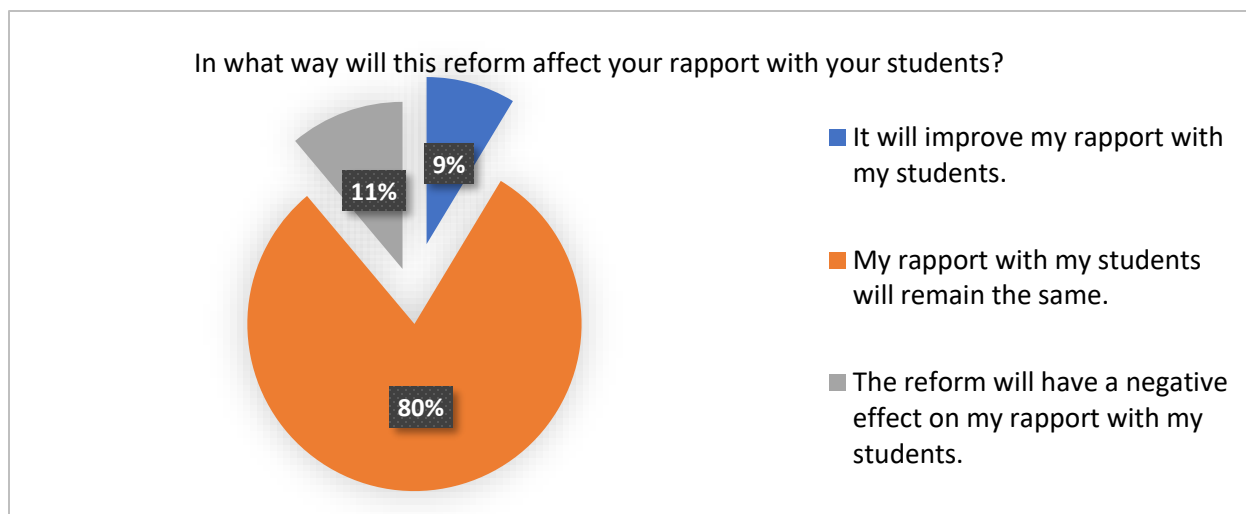


Figure 14: Teachers' perception of how this reform will affect their relationships with students.

The majority of the participating teachers in this study arguably do not share the views of Hargreaves *et al* (2002) who argue that Teacher-Based Assessment is more humanistic and hence leads to better relationships, but neither do they seem to be viewing this as being a matter of surveillance which forces them to extend Foucault's "normalizing gaze" due to being overseers of the SEC assessment. In effect, the majority do not see this as changing their relationships with students. When testing for a relationship between this belief, gender, and qualifications, no significant relationship was found. However, when testing whether teachers' views were related to their years of experience, a significant relationship was found ($p<0.01$). Cramer's V was calculated to test the strength of the relationship between the two variables, and the result was found to be 0.48, indicating a moderate relationship. The implication of this is that years of experience has a bearing on how teachers perceive assessment reforms as impacting their relationships with students.

The researcher posited that the SEC assessment reform will have an effect on teachers' accountability and a question on this revealed that over half the respondents (56%, $n=45$) envisage no change to the levels of accountability that they now experience; however, 42% ($n=34$) feel that the reform will make them more accountable. The divide in teachers' views is quite striking (Figure 15).

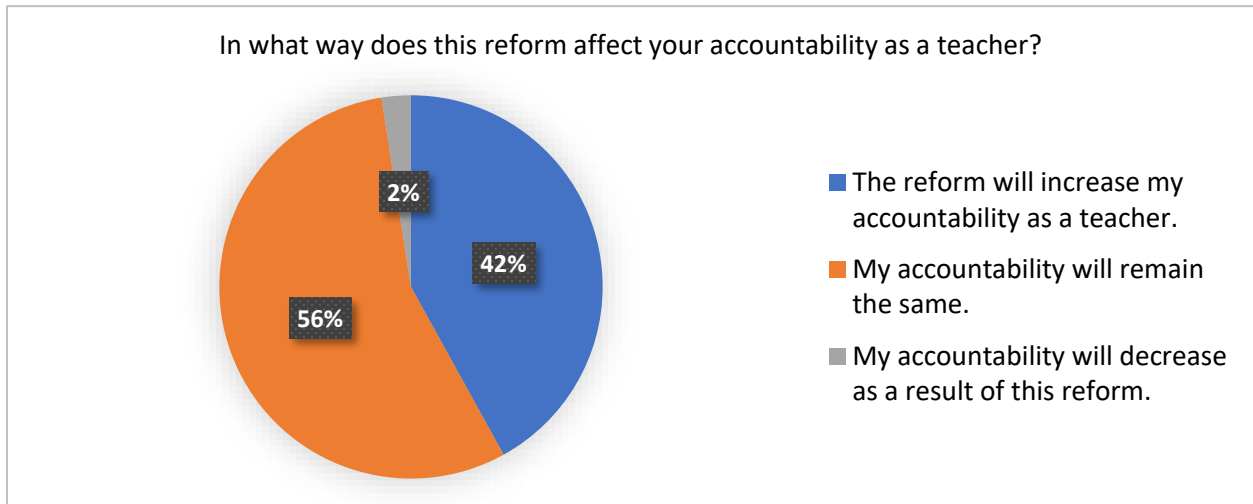


Figure 15: Teachers' perception of how this reform will affect their accountability as teachers.

Brown *et al's* (2009) research shows that when accountability is increased through Teacher-Based Assessment, an increase in direct transmission is also recorded, with this not being in line with *Learning Outcome Framework's* pedagogical philosophy. Further to this, their study also shows that when teachers are made more accountable, they place greater pressure on learners to be accountable for their work. In Brown *et al's* (2009) research, the teachers transferred the accountability measures to the students, by introducing new measures which were not a part of their previous classroom practice and their effects were dubious. In turn, this means that when teachers are held more accountable, these accountability measures are then pushed forward to the learners, implying that the learners will be at the end of the chain when it comes to dealing with the new pressures. Chi-Squared Tests found no significant relationships between teachers' gender, experience, and qualifications, implying that these findings are random. The effect of this perceived increase in accountability is yet to be seen in the Maltese context.

The assessment reform, and the Teacher-Based Assessment aspect of it, will spill out of the schools into the public domain. The survey explored whether teachers had a sense of how the public will receive the change and the increased teacher agency in assessment. Well over half the respondents (57%, $n=46$) think that the general public is not interested in Teacher-Based Assessment and a quarter (25% ($n=20$)) think that the general public will be receptive of it. 18% ($n=15$) think that the general public will reject teacher-based assessments (Figure 16).

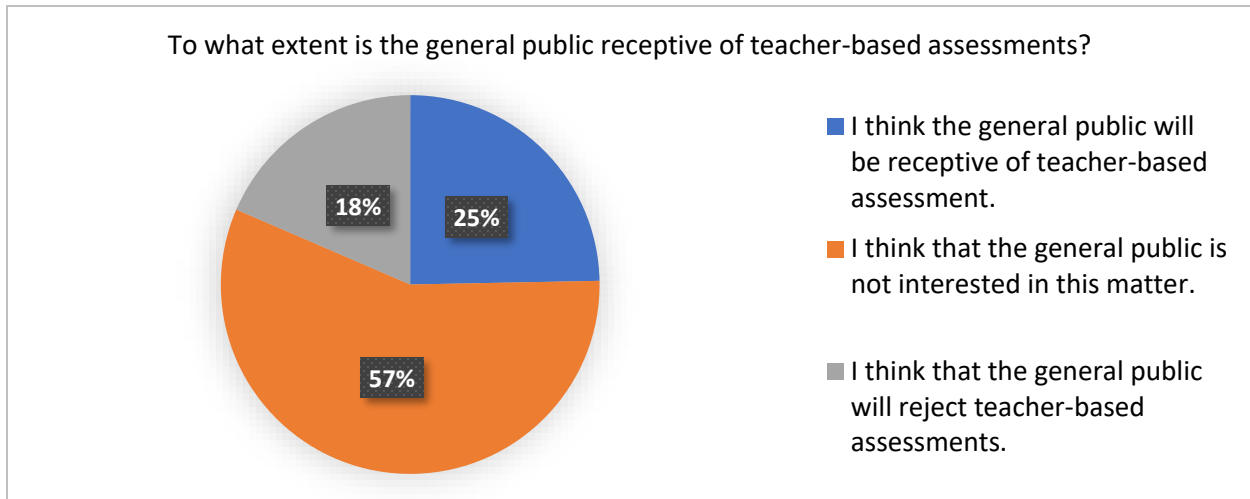


Figure 16: Teachers' perception of how the general public will receive teacher-based assessment.

Whilst the majority of the teachers in the sample do not believe that the general public is interested in this matter, the fact that the reform is not yet in place could possibly have a bearing on this belief. The reviewed literature has shown that in many other contexts, a given society's general public held strong views about this topic (Cheong, 1999; Luyten and Dolkar, 2011; Berry, 2011; Nsibande and Modiba, 2012). There is the possibility that the general public has not yet been given the opportunity to respond to this reform, given that it is still finding its place in the educational system. Future research could possibly investigate this topic when the reform is in place, and when students have begun being assessed via this new system.

The final item for this part of the survey focused on how the participants perceived this reform as affecting their status as teachers. Davison's (2004) research attests to how teachers respond differently to having to assess parts of High-Stakes examinations in class, depending on how their culture views their status. In his study, teachers who viewed themselves as having to compete for status with the results of end-of-school examinations and were found to be deflating students' marks. Most of the participants (89%, $n=72$) think that this reform will not change their status as teachers (Figure 17). Chi-Squared Tests found no significant relationship between the participants' genders, experience, and qualifications for this item.

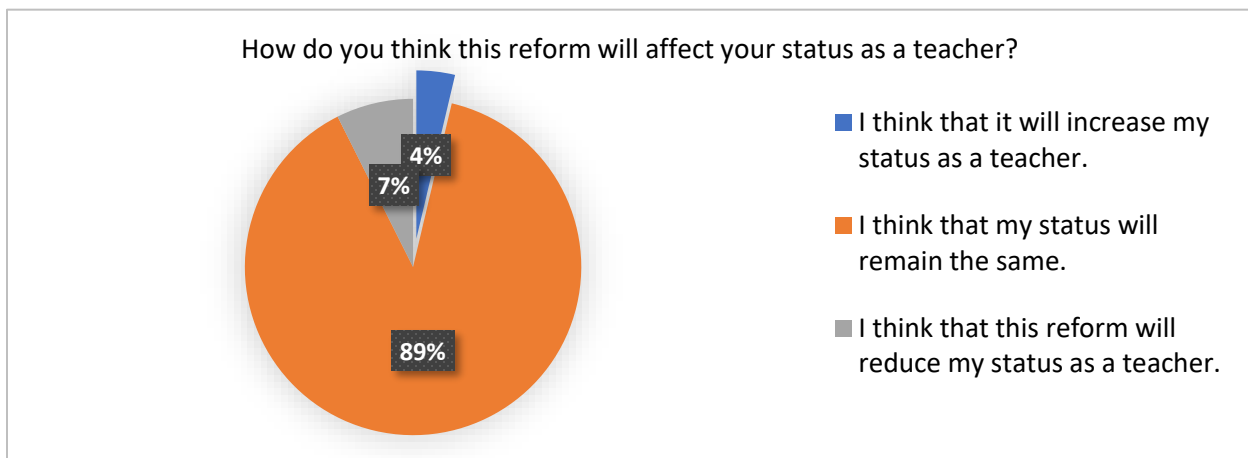


Figure 17: Teachers' perception of how this reform will affect their status as teachers.

Given that accountability and status are both part of how a society views and treats teachers, a further Chi-Squared Test was conducted and the p -value was found to be 0.003, implying that a significant relationship exists between how teachers think that this reform will affect their accountability, as well as how they think it will affect their status as educators. Cramer's V was calculated for these two variables, and the p -value was found to be 0.32, indicating a moderate relationship between the two variables. Effectively, this shows that the participants are conscious of their position in the social system, and whilst they do not perceive a change taking place when it comes to their status, it is worth considering that this may change when their accountability changes, given the significant relationship between the two.

4.12: Teachers' language testing and assessment training and professional development.

Further training is crucial for any reform which impacts teachers' practice. In fact, De Lisle (2016) points out that the biggest cause of unsatisfactory performances by teachers when it comes to Continuous Assessment systems is undertraining. An item was placed in this survey to see what level of training and professional development the sample group had received prior to this reform. As an ethical consideration, it is worth pointing out that training sessions were cancelled by the responsible authorities due to the Covid-19 pandemic, as the participants revealed. The data still shows some interesting findings, however. The participants were asked to either select the option which best described their training in language testing and assessment or add options if the ones provided were not sufficient. The data collected shows that whilst the majority of the participants have received some form of language testing or language assessment training at some point in their careers (either during their initial training, on Postgraduate Courses, or on Professional Development Courses), 24% ($n=19$) chose the option which indicates that they have received no training on language testing or language assessment (Figure 18).

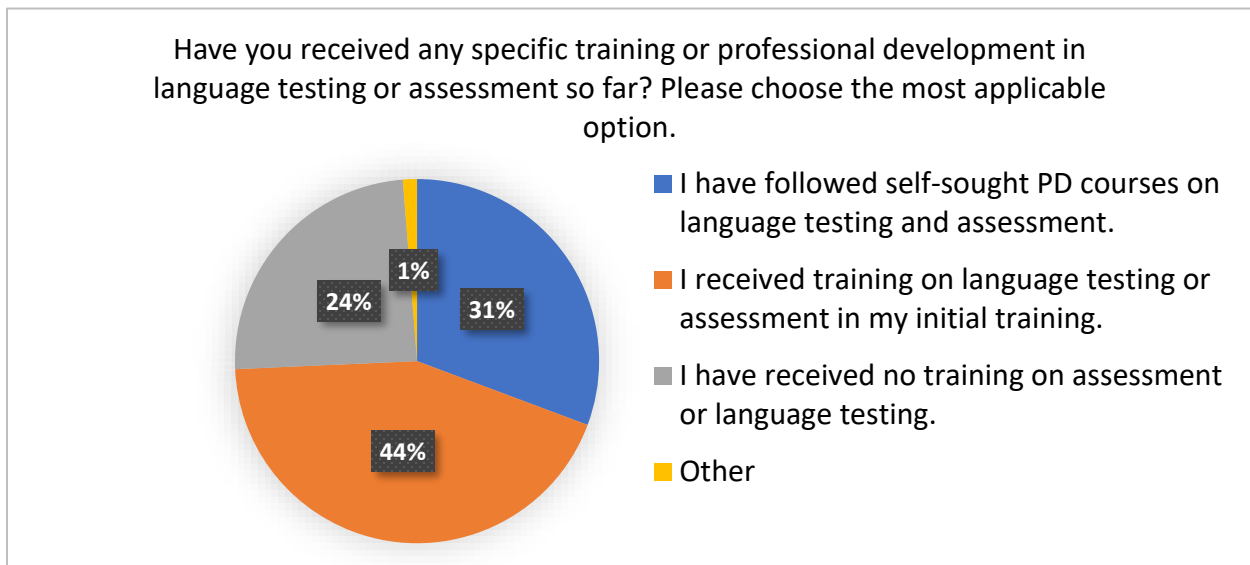


Figure 18: Teachers' language testing and assessment training and professional development.

This is arguably a serious issue given the technical nature of the proposed assessment reform, and also in how research shows that the most crucial component for an educational reform's success is the amount of training which teachers have had. Whilst it is worth considering that the pandemic has disrupted the educational system in general, Cohen *et al*'s (2000) study found that the best professional development is sustained over time rather than being a one-time event, yet one cannot see how this will be possible in the local context.

4.13: Teachers' conceptualization of assessment.

In this section of the survey, the participants were presented with 4 Likert-Scale items containing different statements, each representing different views about assessment types. The participants were asked to mark a number ranging from 1 to 4, with 1 representing strong disagreement, and 4 representing strong agreement. Gilham (2008) points out not providing a neutral option or allowing participants to insert their own responses may seem forceful. Considering that this study aims to find out what the participants' receptivity is and given that bias is an inherent part of receptivity, this 4-point-scale was purposefully used in order not to allow the participants to choose a neutral option.

The first item in this section asked the participants to consider whether they think that Formal Examinations are more important than Classroom Assessments. The majority of the participants (58%, $n=47$) agree or strongly agree with this statement, whilst the rest disagree (Figure 19).

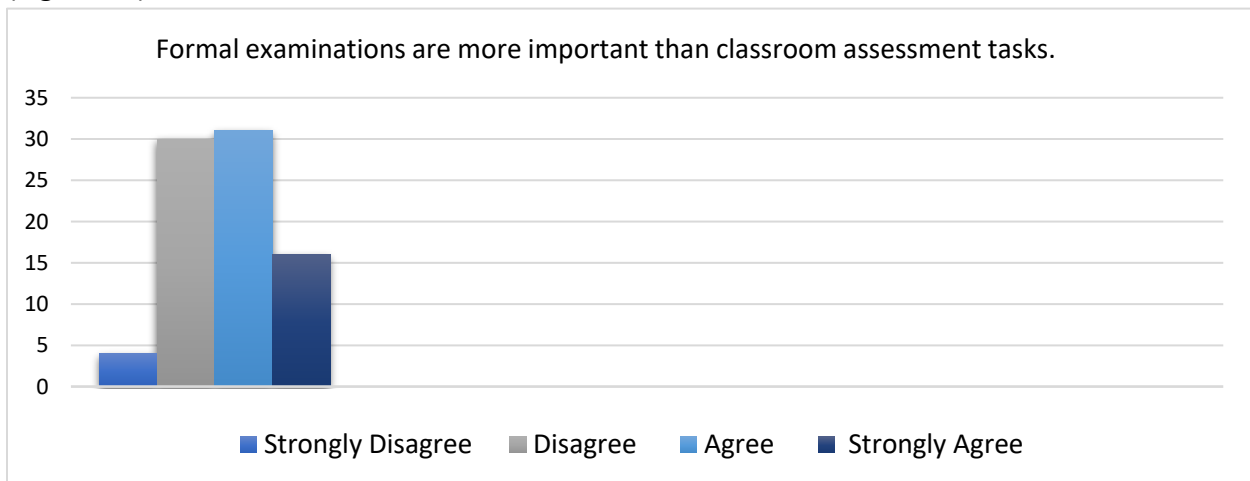


Figure 19: Teachers' conceptualization of assessment 1.

This is reflective of an assessment issue highlighted by Brown (2004), who argues that different teachers have different understandings of what assessment is, and what the associated practice with assessment should be. Given that Classroom Assessment is traditionally associated with enhancing learning, and examinations are typically associated with an end-of-course method of assessing learning, it follows from this item that most of the participants view assessment as being about recording learning, rather than about improving learning.

The following item in this section sought to see whether the participants believe that Classroom Assessments are more important than Formal Examinations. It is worth noting that this statement states the exact opposite of the previous one for reliability purposes. This data shows that 77% ($n=62$) either disagree or strongly disagree with this claim, whilst the rest agree. (Fig. 20).

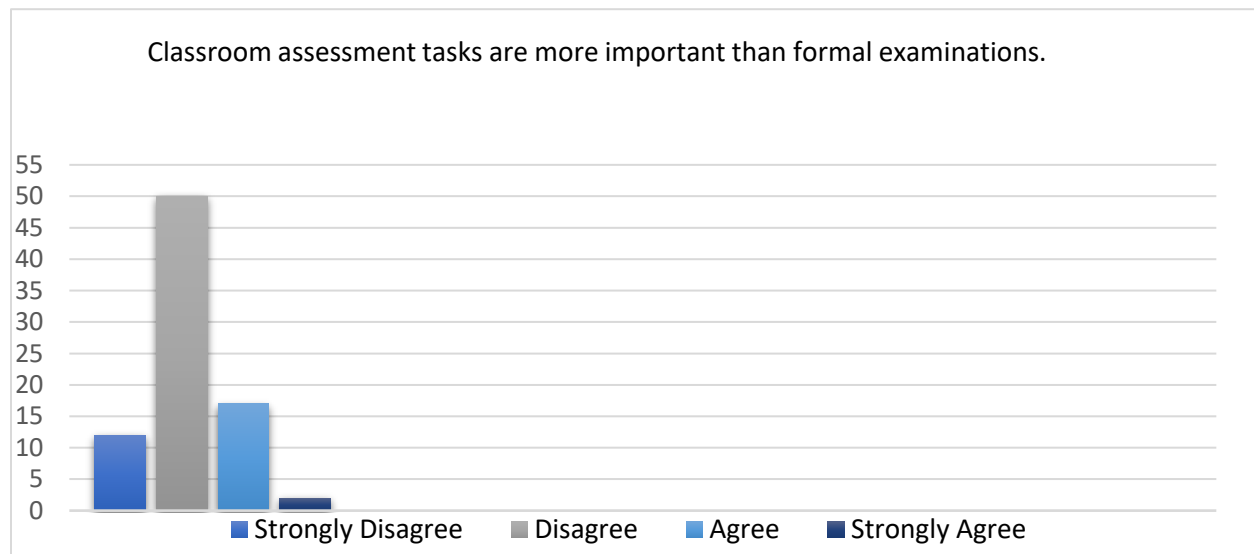


Figure 20: Teachers' conceptualization of assessment 2.

A Chi-Squared Test was conducted using the responses for this item and the previous one, and the p -value was found to be 0.0004, indicating a significant relationship between the two responses. Cramer's V was calculated to test the strength of the relationship between these two variables, and a correlation coefficient of 0.35 was found. This is indicative of a moderate relationship between the two variables in question and attests to the reliability of the responses provided by the participants. Given the existence and strength of this relationship, it

is sensible to argue that the participants in this study share similar beliefs with those in local Church schools, as De Giovanni (2019) points out that the majority of her sample prefer examinations to measure progress over a scholastic year, rather than other types of assessment.

The third item in this section provided the participants with a neutral statement on this topic, by asking them whether they believe that Formal Examinations and Classroom Assessments are equally important. The data show that a strong majority (69%, $n=56$) agree with this statement.

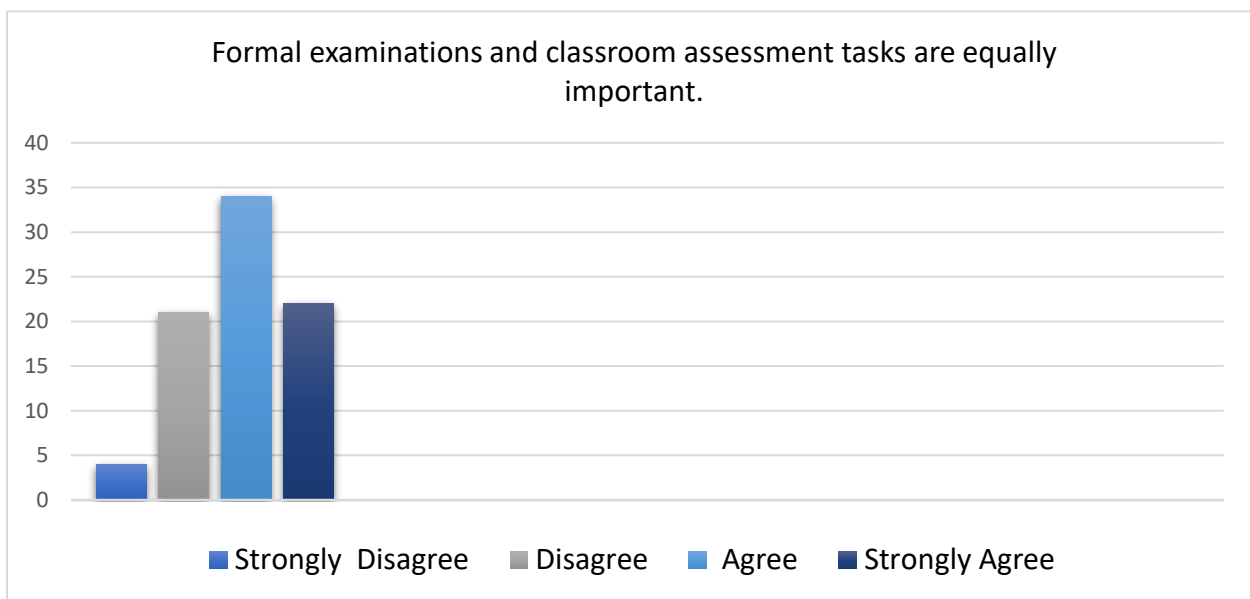


Figure 21: Teachers' conceptualization of assessment 3.

These numbers provide insight into why some of the participants may have disagreed with the previous two statements in the survey, as whilst some of them may have had a preference for a particular assessment type, the majority (69%, $n=56$) agree or strongly agree that both types of assessment methods are of equal importance. This provides insights into how the participants of this survey think about different types of assessment and their value, yet it is worth considering Woods' (1994) and Borg's (2006) arguments that what teachers think, and how they act, may be separate from one another. Further research may be required in the future to see whether the teachers perceive this reform as being coercive.

The final question in this section directly asked the participants whether they think there is a place for School-Based Assessment in the SEC examination. The majority of the participants (65%, $n=53$) disagree or strongly disagree with the statement, whilst the rest agree (Figure 22).

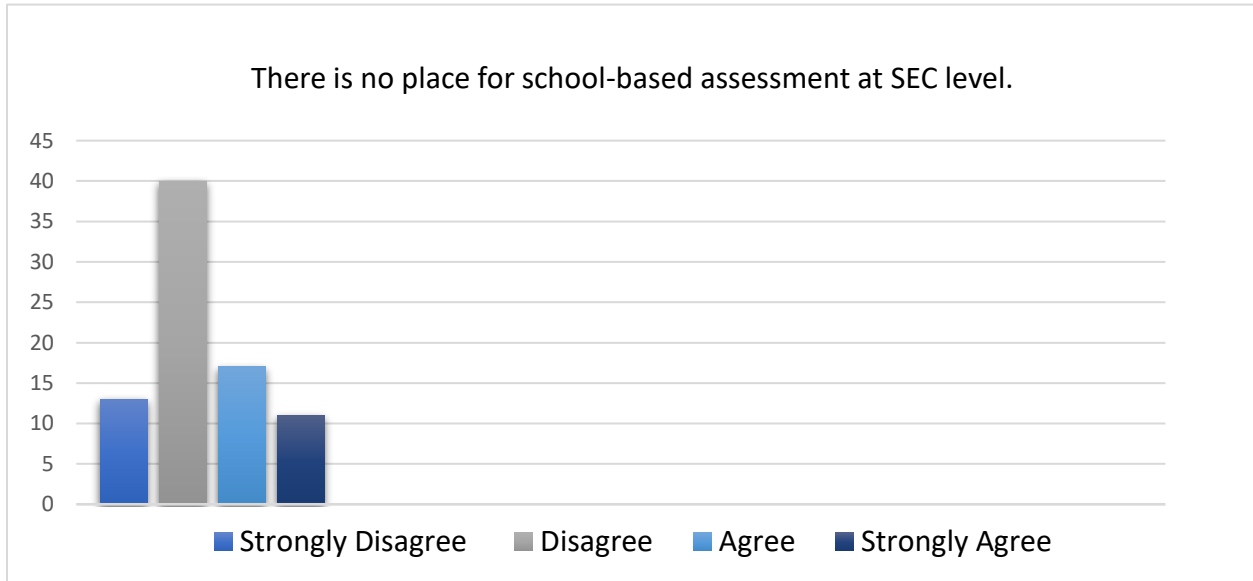


Figure 22: Teachers' conceptualization of assessment 4.

Given this data, it appears that the responses for the first two items were caused by the rejection of radical leanings towards one type of assessment over another. A Chi-Squared Test was conducted using the data from the previous questions and data from this question. Teachers' responses to the item which states "Classroom assessment tasks are more important than Formal Examinations" and the one which states "Formal Examinations and Classroom Assessment tasks are equally important" bore no significant relationship ($p=0.66$, $p=0.42$) with the issue of whether School-Based Assessment should be a part of the SEC. The item which states that "Formal Examinations are more important than Classroom Assessment tasks" was found to have a significant relationship ($p=0.00004$). Cramer's V was then calculated for these two items, revealing a correlation coefficient of 0.38. This shows a moderate relationship between the two variables, implying that the researcher can reliably say that whilst the teachers who responded see Formal Examinations as being more important than Classroom

Assessment tasks, this does not mean that they believe that there is no place for the latter in the new SEC system.

4.14: Teachers' self-efficacy and assessment literacy.

This section of the survey dealt with how efficacious the participants think they are when it comes to assessing parts of their students' SEC qualification. Imbar-Lourie (2008) points out that teachers who are assessment literate can call upon their knowledge and experience both to improve students' learning, as well as conduct their own personal research in order to understand trends in assessment in their working environment. This section took the parts of the Assessment Literacy Skills inventory provided in Giraldo (2018) which the researcher deemed to be the most relevant to this reform and asked the teachers to evaluate their efficacy on each skill. A total of six skills were identified and the seventh item in this section dealt with the issue of washback, which was transposed into a skill in how teachers arguably need to recognize its effects in order not to allow it to be detrimental to students' learning. Hamp-Lyons (1998) points out that excessive washback has detrimental effects on learning, so recognizing its effects is justified as being an assessment skill.

Brookhart (2011) points out that an essential assessment skill for teachers is that they can make use of assessment exercises to improve students' learning. This is also corroborated by Giraldo (2018) and is reflective of Black and William's (1998) arguments that assessment is there to serve learning purposes. The first item in this section asked the participants if they are efficacious when it comes to using assessment tasks to improve their students' learning. Out of 81 surveyed participants, over two thirds (68%, $n=55$) agreed, and a quarter (25%, $n=20$) strongly agreed (Figure 23).

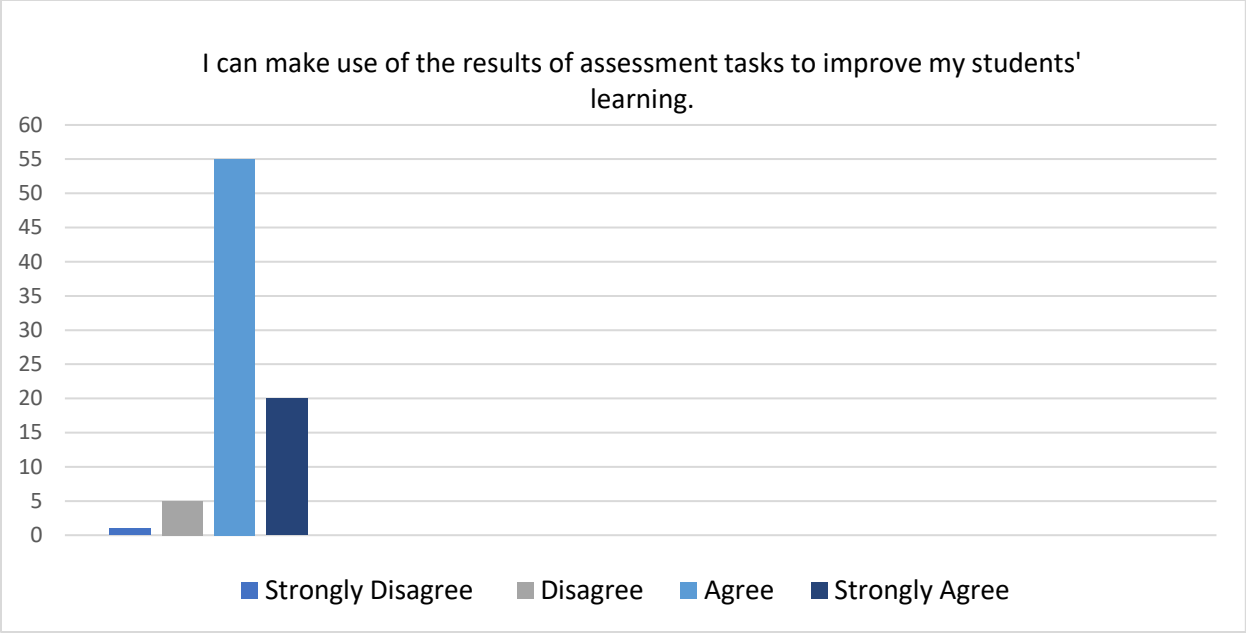


Figure 23: Teachers' self-efficacy when it comes to using assessment tasks to improve students' learning.

This shows that the majority of the participants see themselves as efficacious in their ability to use assessment results to improve their students' learning. Giraldo (2018) considers this a key skill for teachers to have, so it would be worth looking into why 6 participants chose the "disagree" and "strongly disagree" options, given that all learners are entitled to receive the same level of education. Chi-Squared Tests for this item revealed no significant relationship between the participants' genders and experience, yet significance was discovered between their qualifications and responses ($p < 0.01$). Cramer's V was calculated using these variables and a result of 0.68 was gathered, indicating a strong relationship between qualifications and teachers' perceived ability to improve student learning. This data shows that more qualified teachers are more efficacious in their ability to use this skill.

The next item in this section dealt with the teachers' perceived ability to improve their teaching methods as a result of students' results on assessment tasks. Out of 81 respondents, well-over two thirds (69%, $n=56$) agreed, and just under a quarter (23%, $n=19$) strongly agreed. (Fig. 24).

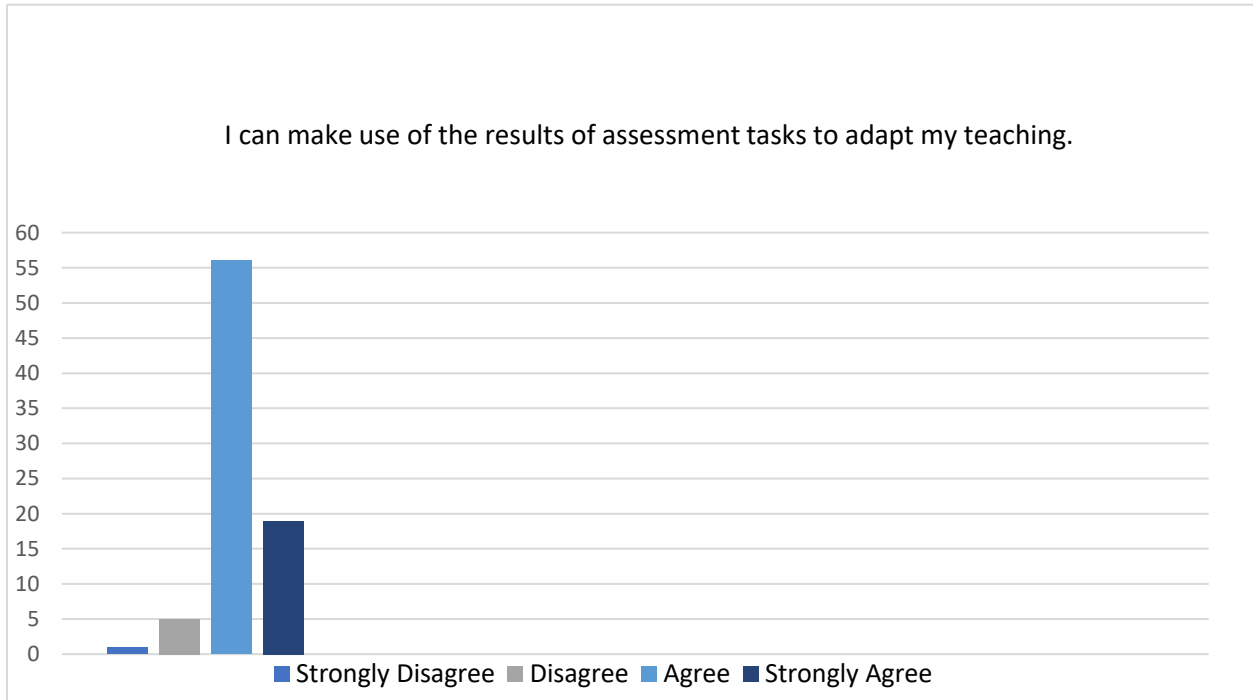


Figure 24: Teachers' self-efficacy when it comes to adapting their teaching as a result of students' results on assessment tasks.

Again, this is indicative of a sample which is efficacious in their ability to make use of assessment tasks to adapt their teaching according to contextual needs. Chi-Squared Tests found no relationship between teachers' genders and experience for this item, yet a significant relationship ($p=0.001$) was found. Cramer's V was calculated again, indicating a moderate relationship through a correlation coefficient of 0.44. This follows the trend noticed for the previous item. A Chi-Squared Test was conducted using the response for the first and second item in this section and the p -value was found to be $<.01$, indicating a high level of dependency between the responses for both questions. Cramer's V was calculated using the responses for these 2 items, and a correlation coefficient of 0.99 was found. This shows that the teachers in this study are, by and large, efficacious in their ability to use students' results to improve their teaching and learning, with the two of these being strongly related to each other.

The survey explored participants' level of efficacy in preparing tasks for students' SEC assessment, should the reform require them to do so. The English Language and Literature syllabi specify the types of tasks which are to be assessed, yet do not prescribe specific tasks which are to be used to assess the students. The two assessment skills are connected to each other in how, when preparing a task, one also needs to consider how the different activities or exercises will be marked. The reform requires teachers to produce a quantitative result, so marking schemes need to be considered.

When it comes to preparing assessment tasks, the majority (70%, $n=56$) claim to be efficacious in this skill, whilst the other 30% ($n=25$) are not. (Fig. 25). One notices a steep

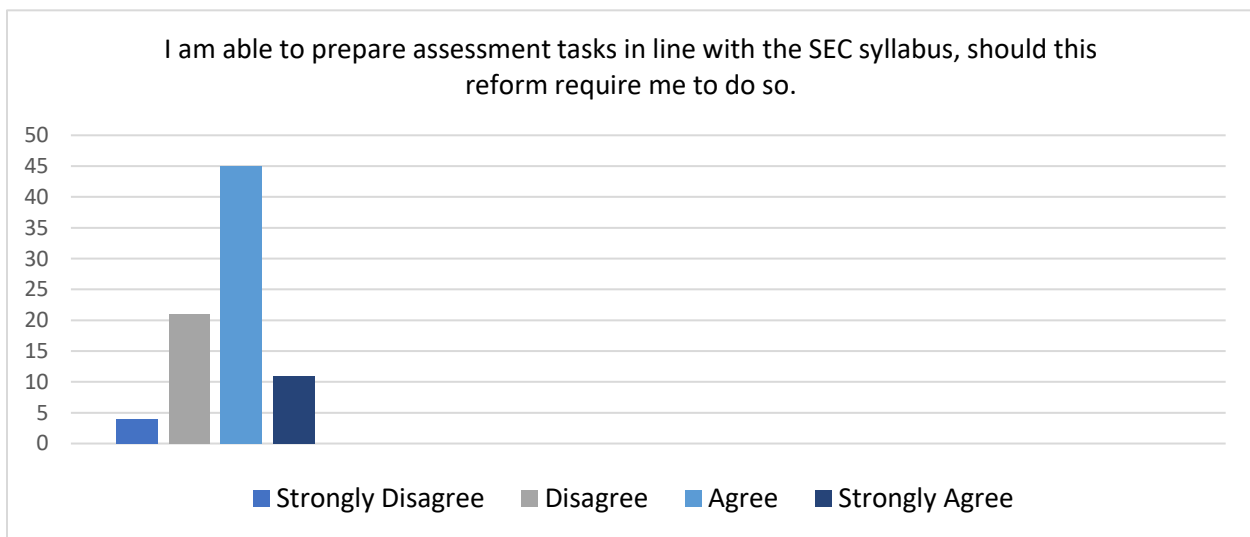


Figure 25: Teachers' self-efficacy when it comes to preparing tasks in line with the SEC syllabus.

Giraldo (2018) highlights how preparing tasks which are reliable and fair for assessment are a crucial part of teachers' assessment literacy. Considering this, these numbers are worrying. Chi-Squared Tests found no significant relationships between gender, experience, and qualifications, meaning that this issue is not particular to a sub-group in the study.

The issue of using rating scales accurately, and reliably, was considered in the next item. A combined 74% of the respondents agree and strongly agree that they are able to use assessment rating scales consistently. , although the fact that there is a minority that is not efficacious is concerning given that they can confound the reliability of any marks provided for students' work (Figure 26).

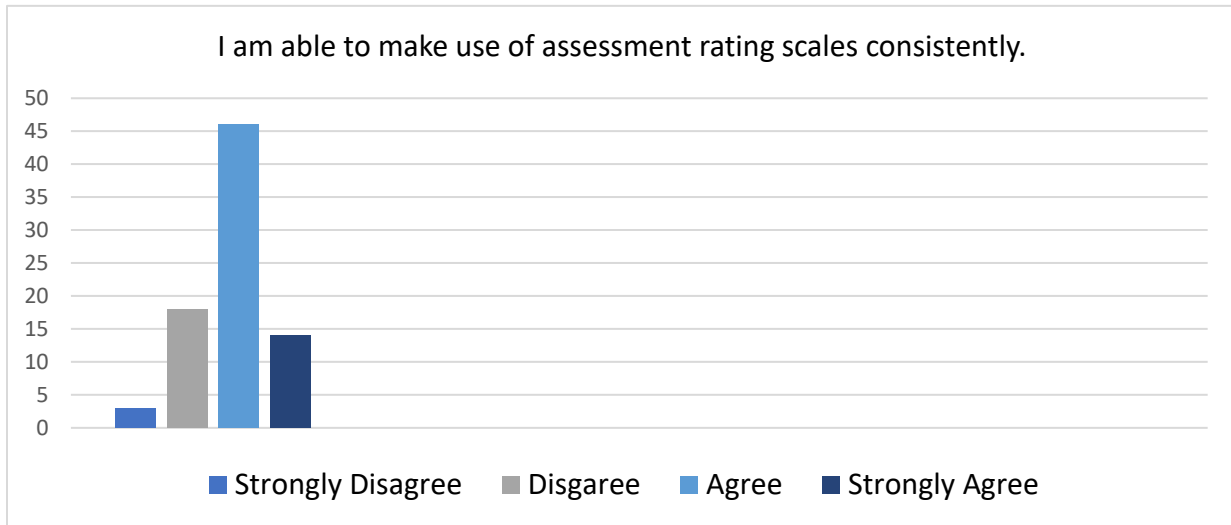


Figure 26: Teachers' self-efficacy when it comes to using rating scales consistently.

A chi-squared test of independence was conducted using the results of for the two items, and a *p-value* of <0.01 was found, indicating a significant relationship between the two. Furthermore, Cramer's *V* was calculated and the correlation coefficient of the two variables was found to be 0.53, indicating a strong relationship. This shows that the self-efficacy required to prepare assessment tasks and the self-efficacy required to use assessment rating scales consistently are related to each other. Teachers with a high-self efficacy for one skill are arguably likely to have high self-efficacy in the other, yet this consistency may be problematic for the teachers who rated themselves as having a low level of self-efficacy when it comes to using these skills.

Finally, this survey revisited the issues of reliability and validity. This time, however, the questions were worded as skills required to evaluate a given assessment task. Giraldo (2018) also recognizes being able to evaluate an externally provided assessment task in terms of its reliability and validity as being a crucial part of teachers' assessment literacy, and given that the majority of the participants would prefer the tasks to be provided by external bodies, it is arguably sensible to consider how efficacious they are in evaluating such tasks. The first question dealt with reliability, whilst the second dealt with validity. The data show that 67% expressed agreement and strong agreement with the statement pertaining to reliability, showing that the majority of the participants seem efficacious in their ability to evaluate how reliable an assessment task is. (Fig. 27). However, the number of participants who do not feel confident when it comes to this skill is higher than that for the previous skills. This may cause issues given that reliability rests on all assessors being consistent.

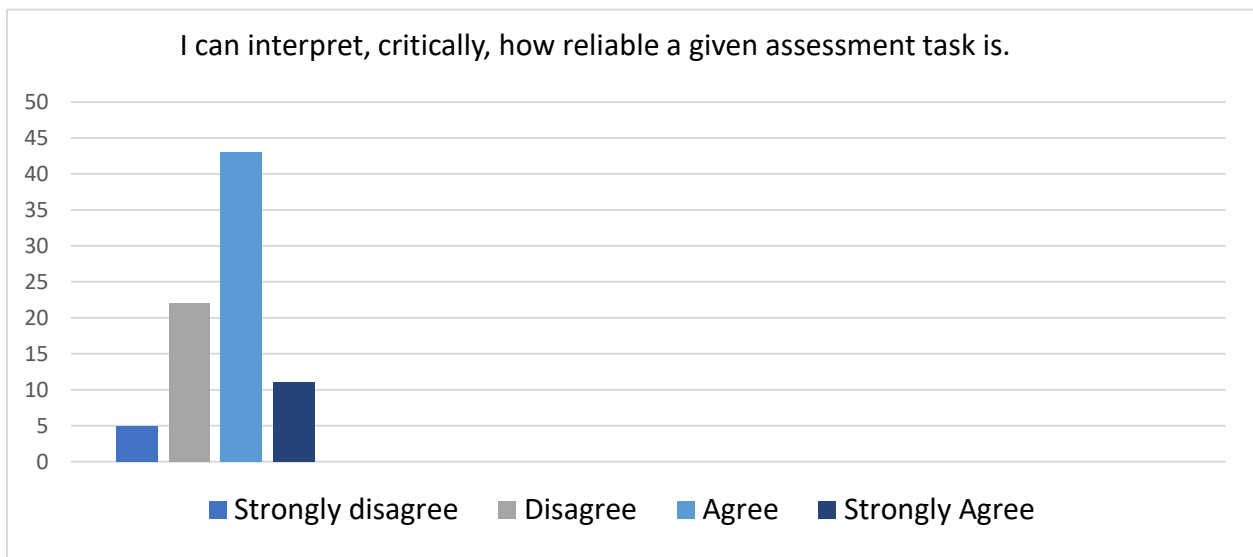


Figure 27: Teachers' self-efficacy when it comes to interpreting how reliable an assessment task is.

Turning to the validity of assessment tasks, a combined 69% reported that they were confident or confident to critically interpreting how valid a given assessment task is. (Fig. 28). Black (2001) points out that validity refers to how well a given task assesses the skills or knowledge which it claims to assess.

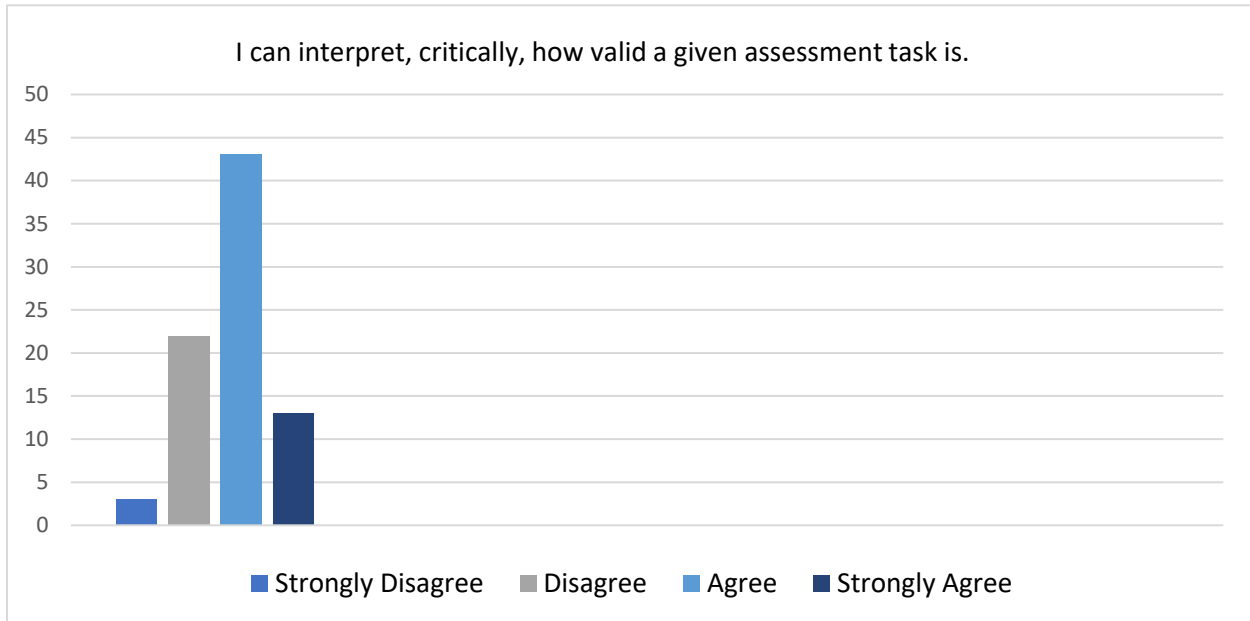


Figure 28: Teachers' self-efficacy when it comes to interpreting how valid an assessment task is.

Whilst the majority agree with the statement, it is concerning that a considerable minority do not, given the importance of validity in assessment. A Chi-Squared Test was conducted using respondents' responses for this item and the previous one, and the *p-value* was found to be <0.01, indicating significance. Cramer's *V* was calculated for this set of variables and the correlation coefficient of 0.77 was found, indicating that a strong relationship exists between the how the participants perceive their self-efficacy when it comes to interpreting both the reliability and validity of given assessment tasks. This follows the arguments of Harlen (2005b), and others, who argue that the two are tightly connected in assessment tasks, yet this study shows that efficaciousness when it comes to the two are also connected.

The final item for the survey dealt with how aware the participants felt they were of the washback effects of assessments. Alderson and Wall (1993) identify two types of washback, positive and negative washback, yet the notion of washback was presented as one item, given

that both positive and negative washback refer to the effects which assessment may have on teaching and learning. A strong majority of 81% ($n=66$) perceive themselves as being knowledgeable of the effects of washback on teaching and learning. (Fig. 29).

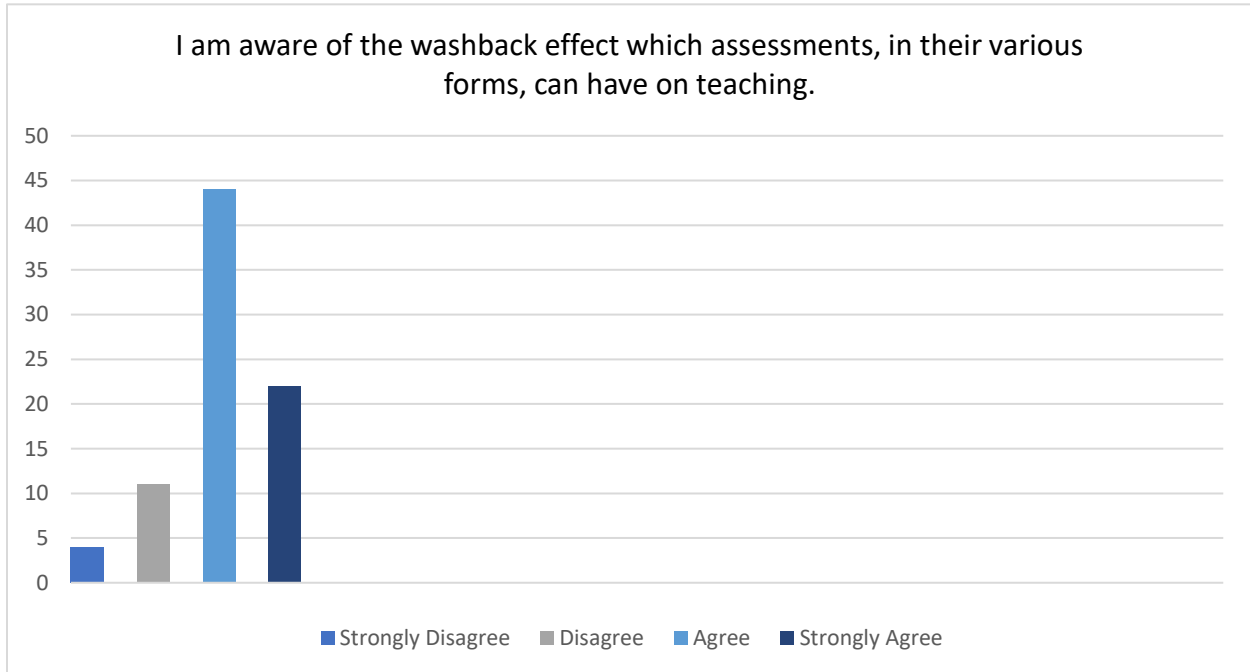


Figure 29: Teachers' self-efficacy when it comes to understanding washback effects.

Given that Alderson and Wall (1993) point out that excessive washback has dubious effects on learning, this is arguably positive. Chi-Squared Tests were conducted between the teachers' years of experience and the response for this item, and the result was found to be $p=0.04$, indicating statistical significance. Cramer's V was calculated for these two variables, and the correlation coefficient was found to be 0.3, indicating a modest relationship. The implication of this would be that more experienced teachers are more efficacious when it comes to recognizing the washback effects of assessment tasks.

In addition, testing for a relationship between awareness of backwash and teachers' qualifications resulted in statistical significance. ($p=0.05$). Cramer's V was then calculated, and a modest relationship (0.37) was established between the two showing that more experienced and more qualified teachers are more knowledgeable about washback.

4.15: Conclusion.

This chapter presented, analysed, and evaluated the findings of the online survey. The questions used in the survey were formulated on the basis of the literature reviewed and reported in Chapter Two and the findings consequently discussed in the light of that literature. When differences were found between the data gathered in the study, and that found in the Literature Review, these were discussed. Chi-Squared Tests and Cramer's V were used throughout to establish significant relationships in the data and test the strength of these relationships. The following chapter will discuss the conclusions drawn from the study, as well as discuss recommendations for policymakers who have a stake in this reform. Recommendations for future researchers who may wish to build upon the findings of this study will also be made.

Chapter 5: Concluding chapter.

5.1: Introduction.

This chapter aims to achieve four goals. The first part will present the conclusions drawn from the online survey in the form of answers to the Research Questions. In the following section, the limitations of the study will be expounded upon. Finally, two sets of recommendations will be made: one set takes the form of implications for stakeholders in this reform, whilst the second set will be recommendations for research which can lead to a greater understanding of the topic of this dissertation.

5.2: Conclusions drawn from the study.

The three secondary Research Questions will each represent a sub-section for this heading, whilst the overarching Research Question will be discussed throughout. The main Research Question for this dissertation was:

“How are Maltese teachers of English responding to having to assess parts of their students’ SEC examination via course work, as of 2020?” (RQ1).

It follows that the teachers could either respond positively or negatively to different aspects of the reform, or they can consider the reform as not bringing about a change in the other aspects. Considering this, the researcher will analyse how well the respondents are responding to the different aspects of this reform for each secondary Research Question.

5.2.1: Secondary research question 1 (RQ2).

To answer the main Research Question of this dissertation, three secondary Research Questions were devised. The first secondary Research Question for this dissertation was as follows:

“Which factors/ variables have a strong relationship with English teachers’ receptivity of the local assessment reform?”

Effectively, this question sought to see which factors played a role which the participants saw as being relevant when it comes to the reform. The variables which fall under this category were: the amount of information provided by the authorities, the participants' level of involvement in the reform, the role of public perception in the reform, and the influence of international high-stakes examinations.

Most of the participants in this study would like more information with regards to the reform, or do not think that they have enough information. Only a minority of the participants think that they have enough. Literature on the topic shows that when teachers do not have enough information, a negative attitude towards the reform in question tends to develop. (Lumadi, 2011). The researcher concludes that this is indicative of a negative reception, as being informed is necessary for teachers to perform their duties well.

Only a minority of the participants had a high level of involvement in this reform. The majority were slightly involved, whilst a considerable amount were not involved at all. Literature shows that not being involved leads to reforms being seen as imposed, rather than participatory. (Fullan, 2003). The upshot of this is that, the majority do not perceive themselves as playing a key role in this reform and this may not be seen as positive, given the literature on how participation leads to greater commitment and success for all involved.

The vast majority of the participants in this study perceive public perception as playing a role in this reform, and only a minority think that public perception is not related to it. Consequently, one can conclude that teachers of English think that public perception plays a role in the assessment reform and that, in light of Fullan's argument on how public involvement is a sign that the public cares, this is a positive indication that the public cares about how students are assessed.

Almost all of the participants think that Malta's below-average performance on International Examinations is playing a role in this reform indicating clearly that despite various criticisms levelled against the way in which these examinations are conducted, the teachers in this study still perceive them to be impactful. The researcher concludes that these examinations

are, from the teachers' perspectives, playing a role in this reform and that the examinations are highly relevant to educational change in the local context.

5.2.2: Secondary research question 2 (RQ3).

The second secondary Research Question dealt with how the teachers in the study perceived this reform to be changing the way in which they practise their profession. Hargreaves (1992) points out that the way in which teachers experience the education system is phenomenologically different to the way in which other stakeholders, such as policymakers and senior officials, experience being in the system. Hence, the following Research Question was developed to examine how English teachers in particular are perceiving this reform as changing their working lives:

“How do English teachers perceive that this change will affect their working lives, as teachers of English in the Maltese educational system?”

The variables that were studied for this question were the widest in the study, given that the aim of the dissertation is to see how it will impact the teachers of English in Secondary Schools.

The data collected lead the researcher to conclude that whilst the absolute majority see this reform as increasing their workload, and that this reform requires changes to their teaching methods, they are less confident that the reform will lead to an increase in students' learning. The majority think that the reform will not lead to any fruitful gain, although there is a considerable minority who think that it will. The statistical tests lead the researcher to conclude that there is also a gendered pattern involved in this trend, as the female teachers are more likely to perceive it as increasing learning than the male ones., with this relationship being statistically significant. In terms of the main Research Question, the researcher concludes that this is indicative of a negative receptivity, as the increase in work is being seen by the participants as not necessarily leading to an increase in students' learning.

The researcher also deduces that there is an issue present in terms of the participants' confidence when it comes to assessing learners in a reliable and valid way. This is particularly pertinent when it comes to being confident in other teachers' ability to assess reliably. The participants also reported a perceived decrease in the credibility of the SEC qualification. In

conclusion, this is indicative of a negative receptivity on the respondents' sides, as reliable and valid assessments, along with credible national qualifications, are both desirable and essential in any educational system.

There are several mixed findings when it comes to certain aspects of how the participants perceive their position as members of Maltese society being impacted as a result of this reform. The researcher concludes that the responses to items in this section of the survey are reflective of the general ambiguity noted in relation to how teachers think their roles will, or will not, change due to this reform. The participants are divided in terms of who should take charge of preparing the assessment tasks for the SEC examinations, as well as when it comes to how this reform will affect their roles as teachers. These two questions are plausibly connected, given that assessing students for summative purposes and teaching lessons are arguably separate skills. Further to this, the data is rather split when it comes to how teachers think this reform will affect their accountability and the degree of interest they think the general public will take in the assessments they conduct. Although there were answers for these items which were more popular than others, the responses did not show as many overwhelming preferences as in the previous items. The only exception to this was seen on the question regarding teachers' perceived relationship with students, as the majority do not think that this reform will impact their relationship. In conclusion, the researcher concludes that the mixed findings of the questions noted here are reflective of the participants' diverse perceptions on how this reform will affect their positions in society. This does not convey a positive or negative receptivity of the reform, but rather, uncovers the diversity of perspectives present in the population.

A further conclusion is being drawn in relation to teachers' language testing and assessment training and their responses to the Likert scales pertaining to their conceptualization of assessment. The research shows that the participants have various levels of training when it comes to language testing and assessment, and also that there is diversity in their conceptualizations of assessment. The statistical tests lead the researcher to conclude that the teachers are receiving the reform in a positive manner, as they do agree that there is a place for teacher-based assessment at SEC level, yet still think that the formal examination

should still have a higher weighting than the classroom assessments. The diversity in conceptualizations of assessments is arguably a result of the different types of training which the participants have, yet one cannot exclude the fact that a considerable amount have no training in this area. The impact of this on the responses could not be established. Finally, the researcher concludes that, in light of arguments made by Bandura on self-efficacy, the lack of training could lead to a negative reception given that the participants will be asked to engage in an activity which they are not trained to do.

5.2.3: Secondary research question 3 (RQ4).

Being skilled at assessing learning is often seen as being separate from teaching, although some point out that the two are merging into each other because of the increasing assessment responsibilities being given to teachers. (Brookhart, 2011). This part of the chapter will present the conclusions regarding how efficacious teachers are when it comes to various assessment skills.

The conclusion drawn for this research question is that whilst the majority are efficacious on each skill, there are certain skills which a minority, of varying sizes depending on the skill, are not efficacious. This is problematic in light of how assessment skills are essential for ethical and credible assessments to take place. Further consideration of Bandura's notion of self-efficacy lead the researcher to conclude that there are a minority who are not receiving this reform positively, as they are not efficacious in the skills which the reform requires them to have.

5.3: Limitations of the study.

Like many other studies, this study also has its limitations. In this part of the chapter, three limitations will be expounded upon: the lack of a qualitative aspect, the relatively low response rate, and how mood may have come to be a confounding variable in the data collected.

One of the greatest limitations of this study was that it relied solely on a quantitative methodology. Teddlie and Tashakkori (2008) point out that when researchers precede a quantitative study with qualitative research, this allows the beginning phase to influence "the

generation of research questions and hypotheses (or both) for the second phase.” (Teddlie and Tashakkori, 2008, p.109). In effect, removing the qualitative aspect led to the researcher imposing his own biases, based on the literature yet biases none the less, on the Research Questions. Had the qualitative aspect taken place, greater validity would have also been achieved through a more grounded theory approach to collecting data and refining research questions.

Issues of sample size and respondents’ mood are also worth considering as limitations for this study, given that they may influence any data gathered from replications of this study. The response rate was that of 32.4%, which is rather low. Had the response rate been higher, generalizations may have been possible. Further to this, given that the study focused on how people think and feel about a topic, mood may have been in issue when it comes to how optimistic or pessimistic the respondents were when answering. Heide and Gronhaug (1991) point out that human emotions have long been a concern for survey researchers, given that these may influence one’s world view. With this in mind, it is possible that the individual moods of participants on a given day may have affected their responses on the survey.

5.4: Recommendations for policymakers.

Given the practical nature of this research, the researcher believes that it would be useful for a set of recommendations to be drawn up for policymakers. The thinking process behind these recommendations is, of course, to provide insights which will allow for a more ethical and just educational system for all involved. The following recommendations are being made:

- Given that this study has shown that not all the English teachers think that they have enough information regarding the reform, it would be beneficial if a repository were to be created containing all the necessary information.
- The teachers in this study perceive the reform as increasing their workload, although being a quantitative study, the researcher can only hypothesise on how it could do this. An investigation as to how this may happen could be beneficial in order to ensure fairer working conditions or reassurances of the value of this increased work, if it is warranted.

- The study has shown that not all the teachers are confident in their assessment literacy skills, nor do they all have the same level of training in this field. Providing longitudinal, hands-on training in which the skills required to implement the reform develop will lead to all students having a teacher who has the necessary skills for fairer assessment to be ensured.

5.5: Recommendations for future research.

As an addendum to this dissertation, the researcher would like to make some recommendations for future research which may build on the work done on this topic in the Maltese context. The major recommendation being made is that of studying the same topic using a qualitative methodology. These are the specific suggestions:

- One may wish to study the topic, using similar questions to the ones in the survey in the appendix, from a qualitative perspective using semi-structured interviews. This would lead to data being more valid, rather than being imposed through reviewed literature and the researcher's judgement.
- Further to this, qualitative research would allow for explanations for certain phenomena. For example, the survey data allows the researcher of this study to conclude that public perception is relevant to the reform, but it does not allow him to explain in what way it is relevant.
- A further study, using any epistemology, conducted after the reform is in place, could be compared to the findings of this study to see if teachers' perception of course work at SEC level will have changed.

5.6: Conclusion.

This dissertation has sought to integrate psychometric and socio-cultural issues in assessment through a study which aims to present the receptivity of both the reform's technical and social aspects. The researcher would like to conclude this dissertation by positing a quote from Shohamy (2001) which reminds many that, whilst educational assessment is a highly technical and scientific subject, one must never forget the impact which these have on human's lives, and the ethical implications which this has:

“...the use of tests to make detrimental decisions for individuals is one of the major sources of power.” (Shohamy, 2001, p.17).

List of references.

- Alderson, C.J., & Wall, D. (1993). 'Does Washback Exist?' Paper Presented at a Symposium on the Educational and Social Impacts of Language Tests, *Language Testing Research Colloquium*, February 1992.
- Attard Tonna, M., & Bugeja, G. (2016). 'A Reflection on the Learning Outcomes Framework Project', *Malta Review of Educational Research*, 10(1), 169-176.
- Bandura, A. (1997a). *Self-Efficacy: The Exercise of Control*. New York: W.H. Freeman and Company.
- Bangert-Downs, R.L., Kulik, J.A., & Kulik, C.C. (1991). 'Effects of Frequent Classroom Testing', *The Journal of Educational Research*, 85(2), 89-99.
- Berry, R. (2011). 'Assessment trends in Hong Kong: seeking to establish formative assessment in an examination culture', *Assessment in Education: Principles, Policy & Practice*, 18(2), 199-211.
- Black, P. (1998). *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*. London and Philadelphia: Falmer Press.
- Black, P. (2001). 'Dreams, Strategies and Systems: Portraits of assessment past, present and future', *Assessment in Education: Principles, Policy & Practice*, 8(1), 65-85.
- Black, P., & William, D. (1998a). 'Assessment and Classroom Learning', *Assessment in Education: Principles, Policy & Practice*, 5:1, 7-74.
- Black, P., & William, D. (1998b). 'Inside the Black Box: Raising Standards Through Classroom Assessment', *Phi Delta Kappan*, 80(2), 139-144, 146-148.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). 'Validity in teachers' summative assessments', *Assessment in Education: Principle, Policy & Practice*, 17(2), 215-232.

Borg, C., Mayo, P., & Raykov, M. (2016). *Adult learning in Malta: Insights into current participation, content and forms of adult learning*. Msida: Faculty of Education, University of Malta.

Borg, S. (2006). *Teacher Cognition and Language Education: Research and Practice*. Norfolk: Continuum.

Borg, S. (2003). 'Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do', *Language Teaching*, 36, 81-109.

Brookhart, S.M. (2011). 'Educational Assessment Knowledge and Skills for Teachers', *Educational Measurement: Issues and Practice*, 30(1), 3-12.

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., Lake, R., & Matters, G. (2009). 'Assessment policy and practice effects of New Zealand and Queensland teachers' conceptions of teaching', *Journal of Education for Teaching*, 35(1), 61-75.

Brown, G.T.L., Lake, R., & Matters, G. (2010). 'Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes', *Teaching and Teacher Education*, 27, 210-220.

Bryman, A. (2007). 'Barriers to Integrating Quantitative and Qualitative Research', *Journal of Mixed Methods Research*, 1(1), pp.8-22.

Bryman, A. (2012) *Social Research Methods*. (4th ed.) Oxford: Oxford University Press.

Buldur, S. & Tatar, N. (2011). 'Development of self-efficacy towards using alternative assessment scale', *Asia Pacific Educational Review*, 12, 485-495.

Cheong, C.C. (1999). 'Public examinations in Hong Kong', *Assessment in Education: Principles, Policy & Practice*, 6(3), 405-417.

Cohen, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2000). 'What Makes Professional Development Effective? Results From a National Sample of Teachers', *American Educational Review Journal*, 38(4), 915-945.

Council of Europe. 2015. *Language Education Policy Profile: Malta*. Strasbourg: Council of Europe.

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.

De Giovanni, J. (2019). 'Perceptions of Assessment among Teachers of English in Church Schools', *Malta Review of Educational Research*, 13(2), 210-236.

De Lisle, J. (2015). 'The promise and reality of formative assessment practice in a continuous assessment scheme: the case of Trinidad and Tobago', *Assessment in Education: Principles, Policy & Practice*. 22(1), 79-103.

De Lisle, J. (2016). 'Unravelling continuous assessment practice: Policy implications for teachers' work and professional learning', *Studies in Educational Evaluation*, 50, 33-45.

Deutskens, E., Ruyter, K.D., Wetzels, M., & Oosterveld, P. (2004) Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study. *Marketing Letters*, 15(1), 21-36.

Directorate for Learning and Assessment Programmes. (2018). *FAQs: New Learning Outcomes and Syllabi*, Directorate for Learning and Assessment Programmes, viewed 04/06/2020. < <https://curriculum.gov.mt/en/Pages/FAQs-New-Syllabi.aspx> >

Directorate for Standards and Quality in Education. (2015). *Learning Outcomes Framework: Educators' Guide for Pedagogy and Assessment: English*. Floriana: Ministry for Education and Employment.

Firestone, W.A. (1998). 'A Tale of Two Tests: tensions in assessment policy', *Assessment in Education, Principles, Policy & Practice*, 5(2), 175-191.

- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. New York: Vintage.
- Fullan, M. (1999). *Change Forces: The Sequel*. London: Falmer Press.
- Fullan, M. (2003). *Change Forces with a Vengeance*. London: Routledge.
- Gillham, B. (2007) *Developing a Questionnaire*. (2nd ed.) London and New York: Continuum.
- Gioka, O. (2009). 'Teacher or examiner? The Tensions between Formative and Summative Assessment in the Case of Science Coursework', *Research in Science Education*, 39(4), 411-428.
- Gipps, C. (1994). 'Socio-Cultural Aspects of Assessment', *Review of Research in Education*, 24, 355-392.
- Giraldo, F. (2018). 'Language Assessment Literacy: Implications for Language Teachers', *Profile*, 20(1), 179-195.
- Goldstein, H. (2004). 'International comparisons of student attainment: some issues arising from the PISA study', *Assessment in Education: Principles, Policy & Practice*, 11(3), 319-330.
- Golombek, P.R. (1998). 'A study of Language Teachers' Personal Practical Knowledge', *TESOL Quarterly*, 32(3), 447-464.
- Gray, D.E. (2004) *Doing Research in the Real World*. London: SAGE.
- Grima, G., Camilleri, R., Chircop, S., & Ventura, F. 2005. *MATSEC: Strengthening an Examination System*. Floriana: Ministry of Education, Youth, and Employment.
- Guskey, T.R. (2002). 'Professional Development and Teacher Change', *Teachers and Teaching: Theory and Practice*, 8(3), 381-391.
- Hamp-Lyons, L. (1998). 'Ethical Test Preparation Practice: The case of the TOEFL', *TESOL, Inc.*, 32(3), 329-337.
- Hargreaves, A. (1994). *Changing Teachers, Changing Times: teachers' work and culture in the postmodern age*. London: Cassell.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). 'Perspectives on Alternative Assessment Reform', *American Educational Research Journal*, 39(1), 69-95.

Harlen, W. (2005a). 'Teachers' summative practices and assessment for learning- tensions and synergies', *Curriculum Journal*, 16(2), 207-223.

Harlen, W. (2005b). 'Trusting teachers' judgement: research evidence on the reliability and validity of teachers' assessment used for summative purposes', *Research Papers in Education*, 20(3), 245-270.

Harlen, W., & James, M. (1997). 'Assessment and Learning: differences and relationships between formative and summative assessment', *Assessment in Education*, 4(3), 365-379.

Harrison, C. (2005). 'Teachers Developing Assessment for Learning: mapping teacher change', *Teacher Development*, 9(2), 255-263.

Hattie, J. (2012). *Visible Learning for Teachers: Maximising Impact on Learning*. 2nd edn. New York: Routledge.

Hunter, L. (2012) 'Challenging the reported disadvantages of e-questionnaires and addressing the methodological issues of online data collection', *Nurse Researcher*, 20(1), 11-20.

Inbar-Louire, O. (2008). 'Constructing a language assessment knowledge base: A focus on language assessment courses', *Language Testing*, 25(3), 385-402.

Isaacs, T. (2014). 'Curriculum and assessment reform gone wrong: the perfect storm in GCSE English', *The Curriculum Journal*, 25(1), 130-147.

Jeong, H. (2013). 'Defining assessment literacy: Is it different for language testers and non-language testers?', *Language Testing*, 30(3), 245-362.

Johnson, R.B. & Onwuegbuzie, A.J. (2004) 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher*, 33(7), 14-26.

Little, J.W. (1993). 'Teachers' Professional Development in a Climate of Educational Reform', *Educational Evaluation and Policy Analysis*, 15(2), 129-151.

Lumadi, W. (2011). 'Continuous Assessment in Schools: Teachers' Bitter Pill to Swallow', *Journal of Educational Research*, 7(1), 27-34.

Luyten, H., & Dolkar, D. (2010). 'School-based assessments in high-stakes examinations in Bhutan: a question of trust? Exploring inconsistencies between external exam scores, school-based assessments, detailed teacher ratings, and student self-ratings', *Educational Research and Evaluation*, 16(5), 421-435.

MATSEC. (2020). *SEC 11 Syllabus: English Language: 2023*, MATSEC, viewed 16/04/2020. <
https://www.um.edu.mt/_data/assets/pdf_file/0004/427540/SEC11.pdf >

MATSEC. (2020). *SEC 12 Syllabus: English Literature: 2023*, MATSEC, viewed 16/04/2020.
<https://www.um.edu.mt/_data/assets/pdf_file/0006/427542/SEC12.pdf ?- >

McKim, C.A. (2017). 'The Value of Mixed Methods Research: A Mixed Methods Study', *Journal of Mixed Methods Research*, 11(2), 202-222.

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.

Mee, C.Y. (1998). 'The Examination Culture and its Impact on Literacy Innovations: The Case of Singapore', *Language and Education*, 12(3), 192-209.

Mendoza, A.A.L., & Arandia, R.B. (2011). 'Language Testing in Colombia: A Call for More Teacher Education and Teacher Training in Language Assessment', *Profile*, 11(2), 55-70.

Ministry for Education and Employment. (2014). *Framework for the Education Strategy for Malta 2014-2024: Sustaining Foundations, Creating Alternatives, Increasing Employability*. Floriana: Ministry for Education and Employment.

Ministry for Education and Employment. (2016). *Progress in International Reading Literacy Study (PIRLS) 2016: National Report Malta 2016*. Floriana: Ministry for Education and Employment.

Ministry of Education and Employment. 1999. *National Minimum Curriculum*, Ministry for Education and Employment, viewed 16/04/2020.

<https://education.gov.mt/en/resources/Documents/Policy%20Documents/national%20minimum%20curriculum_english.pdf >

Ministry of Education and Employment. 2012. *A National Curriculum Framework for All*. Floriana: Ministry of Education and Employment.

Morgan, D.L. (2007). 'Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods', *Journal of Mixed Methods Research*, 1(1), 48-76.

Moroz, R., & Waugh, R.F. (2000). 'Teacher receptivity to system-wide educational change', *Journal of Educational Administration*, 38(2), 159-17.

Heide, M. & Gronhaug, K. (1991). *Respondents' Moods As a Biasing Factor in Surveys: an Experimental Study*, The Association for Consumer Research, viewed 04/06/2020. <

<https://www.acrwebsite.org/volumes/7218/volumes/v18/NA-18> >

Moss, P.A. (1996). 'Enlarging the Dialogue in Educational Measurement: Voices from Interpretive Research Traditions', *Educational Researcher*, 25(1), 20-28+43.

Newton, P.E. (2007). 'Clarifying the purposes of educational assessment', *Assessment in Education*, 14(2), 149-170.

Nsibande, R.N., & Modiba, M.M. (2012). '*I just do as expected*'. Teachers' implementation of Continuous Assessment and challenges to curriculum literacy.' *Research Papers in Education*, 27(5), 629-645.

OECD. (2018). *PISA 2015: Results in Focus*. Published online.

Pajares, M.F. (1992). 'Teachers' Beliefs and Educational Research: Cleaning up a Messy Construct', *Review of Educational Research*, 62(3), 307-332.

Ravitch, S.A. (2020). *The Best Laid Plans... Qualitative Research Design During COVID-19*, Social Science Space, viewed on 04/06/2020. < <https://socialsciencespace.com/2020/03/the-best-laid-plans-qualitative-research-design-during-covid-19/> >

Reid, L. (2007). 'Teachers talking about writing assessment: valuable professional learning?', *Improving Schools*, 10(2), 132-149.

Rotin, J. (2003). 'Changing Practice In Dealing With The Sec Examination Course Work Through Action Research', *Journal of Maltese Education Research*, 2(1), 13-25.

Samson, D. & Marongwe, A.D. (2013). 'Continuous Assessment Results versus End-of-Year Examination Marks in Grade 10 Mathematics in Namibia: The statistics and teachers' opinion', *African Journal of Research in Mathematics, Science and Technology Education*, 17(3), 196-205.

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.

Shohamy, E. 2001. *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Harlow: Pearson.

Smith, K. (2011). 'Professional development of teachers- A prerequisite for AfL to be successfully implemented in the classroom', *Studies in Educational Evaluation*, 37, 55-61.

Spiteri, D. (2019). 'English for All: Repositioning English Across the Curriculum', *Malta Review of Educational Research*, 13(2), 194-208.

Spolsky, B. (2008). 'Language Testing at 25: Maturity and responsibility?', *Language Testing*, 25(3), 297-305.

- Tavcar Krajnc, M. (2006). 'External and internal assessment in the final examination in secondary schools in Slovenia', *International Studies in Sociology of Education*, 16(2), 121-137.
- Taylor, L. (2009). 'Developing Assessment Literacy', *Annual Review of Applied Linguistics*, 29, 21-36.
- Teasdale, A., & Leung, C. (2000). 'Teacher assessment and psychometric theory: a case of paradigm crossing?', *Language Testing*, 17(2), 163-184.
- Teddlie, C. & Tashakkori, A. 2008. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioural Sciences*. London: SAGE.
- van den Berg, R. (2002). 'Teachers' Meanings regarding Educational Practice', *Review of Educational Research*, 72(4), 577-625.
- Van Selm, M. & Jankowski, N.W. (2006). 'Conducting Online Surveys', *Quality & Quantity*, 40, 435-456.
- Waring, M. (2017) 'Finding your theoretical position.' In R.J. Coe (Ed.) *Research Methods and Methodologies in Education*, pp. 15-22. Los Angeles: SAGE.
- Waugh, R., & Godfrey, J. (1993). 'Teacher Receptivity to System-Wide Change in the Implementation Stage', *British Educational Research Journal*, 19(5), 565-578.
- Whetton, C. (2009). 'A brief history of testing time: national curriculum assessment in England 1989-2008', *Educational Research*, 51(2), 137-159.
- Willingham, W.W., Pollack, J.M., & Lewis, C. (2002). 'Grades and Test Scores: Accounting for Observed Differences', *Journal of Educational Measurement*, 39(1), 1-37.
- Woods, D. (1994). *Teacher Cognition in Language Teaching: Beliefs, decision-making and classroom practice*. Cambridge: Cambridge University Press.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). 'The centrality of teachers' judgement practice in assessment: a study of standards in moderation', *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.

Zhang, Z., & Burry-Stock, J.A. (2003). 'Classroom Assessment Practice and Teachers' Self-Perceived Assessment Skills', *Applied Measurement in Education*, 16(4), 323-342.

Appendix A: Survey questions.

Survey Questions (Teachers).

Dear Teacher of English,

I am Terry Zammit, a student reading for an M.T.L (Master in Teaching and Learning) degree at the University of Malta. As part of this course, I need to conduct a research study in the form of a dissertation. My dissertation supervisor is Professor Doreen Spiteri.

The title of my dissertation is *English Teachers' Reception of the New Assessment System*. In this study, I will be examining how English teachers are responding to soon having to assess parts of their students' **SEC examinations** in class throughout **Years 9 to 11**.

To achieve this, I would like to cordially invite you to participate in an online survey which will take approximately ten to fifteen minutes to complete.

Participation is voluntary and no inducements are being offered. Should you choose to participate in this online survey, consent will be assumed as soon as you submit. You are kindly requested not to write down your name, or the name of your school, on any part of this survey. You will not be able to withdraw from this part of the study as I have no means of recognizing which submission will be yours.

If you agree to participate, kindly proceed to the rest of the survey. **Please note that even if you teach in a Middle School, you have still been invited to participate due to the possibility of transfers to Secondary Schools in the near future. Your participation is equally important in light of this.**

If you require further information, please do not hesitate to contact me or my supervisor.

Your sincerely,

Terry Zammit

Supervisor's Details:

Mobile number: #####

Name: Professor Doreen Spiteri

Email address: terry.zammit.15@um.edu.mt

Email address: doreen.spiteri.@um.edu.mt

1. Kindly tick your gender: Male Female Prefer not to say.
2. Kindly tick the amount of years you have been teaching English for:
 - 0-5 years
 - 6-10 years
 - 11-15 years
 - 16-20 years
 - 21 years or over

3. Kindly tick the box which best describes your qualifications:
- Lower qualification than a Bachelor's Degree.
 - Bachelor's Degree in Education.
 - Bachelor's Degree in Education + Postgraduate Qualifications.
 - Bachelor's Degree (not in Education) + a P.G.C.E.
 - Bachelor's Degree in my subject area (not in Education) + a P.G.C.E + further Postgraduate Qualifications.
 - Bachelor's Degree (not in Education) + an M.T.L.
 - Bachelor's Degree (not in Education) + an M.T.L. + further Postgraduate Qualifications.
 - Any University Degree, but no formal qualifications in Education.
4. How do you feel about the amount of **information related to your new responsibilities** you have received with regards to this assessment reform?
- I have received all of the information I need.
 - I have received enough information, but would like more.
 - I have not received enough information.
5. Which one of the following best describes your **level of involvement** in the recent assessment reform in Malta?
- I was highly involved.
 - I was slightly involved.
 - I was not involved at all.
6. Which role do you think **public perception** played in this reform?
- A significant role.
 - A minor role.
 - No role at all.
7. What influence do you think Malta's below average scores on **international tests** (such as PIRLS and PISA) have had on this reform?
- A strong influence.
 - A medium-level influence.
 - A weak influence.
 - No influence at all.
8. How do you think this reform will affect your **overall workload?**
- It will increase my workload.
 - My workload will remain the same.
 - It will decrease my workload.

9. Do you think that, as a teacher of English, your workload will be the same as teachers of other subjects? Yes/No.

If yes, kindly elaborate. _____

10. How do you perceive the assessment reform as changing your current **teaching methods?**

- It will require significant changes to my current teaching methods.
- It will require minor changes to my current teaching methods.
- It will require no changes to my current teaching methods.

11. How do you see this reform as affecting your **students' learning?**

- The reform will increase their learning significantly.
- The reform will increase their learning moderately.
- The reform will not increase their learning.
- The reform will harm their learning.

12. How confident are you that **your assessment** of the students' performance on the school-based tasks is **reliable** and will therefore be a true picture of the students' abilities?

- I'm highly confident.
- I'm moderately confident.
- I'm not confident at all.

13. How confident are you that the SEC school-based tasks will be designed in such a way that they faithfully assess the Learning Outcomes?

- I'm highly confident.
- I'm moderately confident.
- I'm not confident at all.

14. How confident are you that **other teachers' assessment** of the students' performance on the school-based tasks is reliable?

- I'm highly confident.
- I'm moderately confident.

I'm not confident at all.

15. How do you think the that the **classroom-based assessment tasks** would best be prepared?

Tasks should be prepared and provided by a central authority.

Individual teachers should prepare their own assessment tasks.

Groups of teachers in the same school should prepare assessment tasks together.

16. Do you think this assessment reform will change **your role as a teacher**?

Yes.

No.

17. In what way will this reform affect your **rapport with your students**?

It will improve my rapport with my students.

My rapport with my students will remain the same.

The reform will have a negative effect on my rapport with my students.

18. In what way does this reform affect your **accountability as a teacher**?

The reform will increase my accountability.

My accountability will remain the same.

My accountability will decrease as a result of this reform.

19. To what extent is the **general republic** receptive of teacher-based assessments?

I think that the general public will be receptive of teacher-based assessment.

I think that the general public is not interested in this matter.

I think that the general public will reject teacher-based assessment.

20. How do you think that this reform will affect your **status as a teacher**?

I think that it will increase my status as a teacher.

I think that my status will remain the same.

I think that this reform will reduce my status as a teacher.

21. How will this reform affect people's trust in the **SEC qualification**?

It will now be more credible as a qualification.

It will retain the same credibility it has had in recent years.

It will lose some of its credibility.

It will lose its credibility completely.

22. The following scales are about **different types of assessment.** Please circle as follows: 1= strongly disagree, 2= disagree, 3= agree, 4= strongly agree.

Formal Examinations are more important than Classroom Assessment Tasks. 1 2 3 4

Classroom Assessment Tasks are more important than Formal Examinations. 1 2 3 4

Formal Examinations and Classroom Assessment Tasks are equally important. 1 2 3 4

There is no place for School-Based Assessment at SEC level. 1 2 3 4

23. Have you received any specific training or professional development in **language testing or assessment so far?** Please choose the most applicable option.

I have followed mandatory PD courses on assessment or language testing.

I have followed self-sought PD courses on assessment or language testing.

I received training on assessment or language testing during my initial training.

I have received no training on assessment or language testing.

Other (please specify): _____

24. The following statements pertain to your **self-efficacy** when it comes to **assessment literacy** when conducting **Classroom Assessments** for students' **SEC examinations.**

Please circle as follows: 1- strongly disagree, 2- disagree, 3- agree, 4- strongly disagree.

a. I can make use of the results of assessment tasks to improve my students' learning. 1 2 3 4

b. I can make use of the results of assessment tasks to adapt my teaching. 1 2 3 4

c. I am able to prepare assessment tasks in line with the SEC syllabus, should this reform require me to do so. 1 2 3 4

d. I can follow an assessment rating scales consistently. 1 2 3 4

e. I can interpret, critically, how reliable a given assessment task is. 1 2 3 4

f. I can interpret, critically, how valid a given assessment task is. 1 2 3 4

g. I am aware of the washback effects which assessments, in their various forms, can have on my teaching. 1 2 3 4

Appendix B: Interview questions. (Not used).

Interview Questions (Teachers)

1. How would you describe your level of involvement in this reform? What has been happening in recent weeks?
2. Do you think that public perception is playing a role in this reform? If so, how? If not, why?
3. Do you think that Malta's low scores on international tests, such as PISA and PIRLS, played a role in this reform?
4. Will these reforms require you to change your classroom practice? If so, how?
5. Do you think that this system is fairer on students? If so, how?
6. Do you think that this assessment reform intends to improve the learning aspect of assessment, or is its focus solely on summative assessment? Why do you think this?
7. Do you think that this reform could pose issues related to reliability in the classroom assessments?
8. ...and what about validity?
9. Do you think that this reform will change your role in the classroom? If so, how?
10. Do you believe that teacher-accountability will play a role in this reform? If so, how?
11. Do you think that our culture in general is receptive of teacher-based assessments?
12. Do you think that the reform of school-based assessment will raise the profile of classroom assessment or the examination? How so?
13. What effect, if any, do you think the reform will have on how students view their learning now that Classroom Assessment will be a part of their examination? How do you think that parents will respond to this?
14. Have the consultation meetings which have been provided thus far provided you with enough information regarding this reform? Please elaborate.
15. Have you received specific training to deal with the reform at SEC level? Do you have any expectations regarding training for this reform? Please elaborate.
16. Do you think that the training which has been provided for the purposes of assessment in Years 7 and 8 is meant to spill over into the classroom assessment of the SEC examination?
17. How would you comment on this reform in terms of time?
18. Could you provide a general comment on how you feel about this reform?

Appendix C: Permission letters.

PERMISSION LETTER- HEAD OF SCHOOL

12/05/19

Dear Head of School,

I am Terry Zammit, a student reading for an M.T.L (Master in Teaching and Learning) degree at the University of Malta. As part of this course, I need to conduct a research study in the form of a dissertation. My dissertation supervisor is Professor Doreen Spiteri.

The title of my dissertation is *English Teachers' Reception of the New Assessment System*. In order to complete this study, I will be investigating the attitudes of English teachers towards the introduction of this new assessment system, with this being done via interviews. I would like to interview one teacher from your school. Should multiple teachers apply, one teacher will be chosen at random. The interviews will last approximately thirty to forty-five minutes. The questions present in the interview will ask the English teachers about what they think the advantages and disadvantages of such an assessment system are, how they think this system will affect their working life, and how it will affect their relationships with other stakeholders within the educational sector.

I would be grateful if you would give me permission to conduct my research study at your school.

Should you give me permission, I would appreciate it should you forward the attached information letter to the teachers of English in your school, so that they may contact me should they wish to participate. Should you give me permission, and should the teachers choose to participate, I would also like to conduct the interviews on the school's premises, during the teachers' free lessons.

Participation is voluntary and participants will suffer no negative consequences should they choose not to participate. No inducements are being offered. The data collected from the interviews will be coded so as to protect the identity of the teachers in question. The collected data will be transcribed and kept in safe storage, whilst the recordings will be destroyed. Upon the publication of my results (June/July, 2020) the transcriptions will also be destroyed.

If you require further information, please do not hesitate to contact me or my supervisor.

Your sincerely,

Terry Zammit

Mobile number: #####

Email address: terry.zammit.15@um.edu.mt

Supervisor's Details:

Name: Prof. Doreen Spiteri

Email address: doreen.spiteri.@um.edu.mt

PERMISSION LETTER- ONLINE SURVEY- EDUCATION OFFICER

Dear Education Officer of English,

I am Terry Zammit, a student reading for an M.T.L (Master in Teaching and Learning) degree at the University of Malta. As part of this course, I need to conduct a research study in the form of a dissertation. My dissertation supervisor is Professor Doreen Spiteri.

The title of my dissertation is *English Teachers' Reception of the New Assessment System*. In order to complete this study, I will be investigating the attitudes of English teachers towards the introduction of this new assessment system. In order to achieve this, I am requesting permission from you to act as an intermediary/ gatekeeper in this research. Should you agree to this, I would be very grateful if you would forward the link to my online survey to all the teachers of English working in the state sector. The survey takes approximately twenty minutes to complete, and can be completed at the teachers' convenience, considering that it is based online. The questions in the survey ask about the advantages and disadvantages of the new assessment system, how the teachers think that this change will affect their working life, and how they think the new system will affect their relationships with other stakeholders in the Maltese Educational sector.

A link to this survey has been attached to the content of the email, so I would be very grateful should you forward it to the teachers of English working in the State Sector. An Information Letter for the teachers has also been attached to the email.

Participation in the survey is voluntary and no inducements are being offered. The data collected from the surveys will be completely anonymous, as I have no means of discovering the IP addresses of the respondents.

If you require further information, please do not hesitate to contact me or my supervisor.

Your sincerely,

Terry Zammit

Supervisor's Details:

Mobile number:#####

Name: Prof. Doreen Spiteri

Email address: terry.zammit.15@um.edu.mt

Email address: doreen.spiteri.@um.edu.mt

Appendix D: Information letter. (Not used).

PARTICIPANT INFORMATION SHEET- TEACHERS OF ENGLISH (INTERVIEW).

Dear Teacher of English,

I am Terry Zammit and I am a trainee teacher of English, reading for an M.T.L (Master in Teaching and Learning) degree at the University of Malta. As part of my course, I am required to conduct a research study in the form of a dissertation. My supervisor is Professor Doreen Spiteri.

The title of my dissertation is *English Teachers' Reception of the New Assessment System*. In this study, I will be examining how English teachers are responding to soon having to assess parts of their students' SEC examinations in class throughout Years 9 to 11. The focus of this study will be on the said teachers' attitude towards this change, on the basis of the advantages and disadvantages of SBA, the ease with which English may be assessed through this system, and how this system will affect the teachers' daily working lives.

In order to achieve this, I would like to cordially invite you to participate in a semi-structured interview which will last in between thirty and forty-five minutes. During this interview, I will be asking you what you think are the advantages and disadvantages of this new system, the ease with which you feel you can assess English using this system, as well as questions on how you feel that this change will affect your working life and your relationships with other stakeholders in the educational sector. I would like to interview one teacher from your college, so if any more should apply, the participant will be selected at random.

Participation is voluntary and no inducements are being offered. You may choose to withdraw from this study at any point, without there being any negative consequences. I would appreciate it should you allow me to audio-record this interview; however, should you not consent to this, I will understand and I will take down notes instead. The data collected from this interview will be analysed for the purposes of my dissertation. Should you allow me to audio-record the interview, a copy of the transcript will be provided for you to read prior to it being used in the dissertation. Any audio-recorded data will be transcribed immediately after the interview, with the audio-recording subsequently being destroyed. The transcriptions will be placed in safe storage, where only I will have access to them. Furthermore, the transcribed recordings will be destroyed upon the publication of my results. (June/ July, 2020).

If you require further information, please do not hesitate to contact me or my supervisor.

Your sincerely,

Terry Zammit

Mobile Number: #####

Email address: terry.zammit.15@um.edu.mt

Supervisor's Details:

Name: Professor Doreen Spiteri

Email address: doreen.spiteri@um.edu.mt