

To Trust or Not to Trust? School-Based Assessment in Physics High-Stakes Examinations in Malta

Eric Zahra

eric.zahra.11@um.edu.mt

Josette Farrugia

University of Malta

josette.farruga@um.edu.mt

Abstract

The Learning Outcomes Framework in Malta seeks to reform the national assessment policy through a collective effort to change the assessment culture in schools (Attard Tonna & Bugeja, 2016). The new Secondary Education Certificate physics examination will consist of one written paper which will carry 70% of the marks while coursework assessed by teachers will carry the remaining 30%. This implies a doubling of the mark for coursework from the current 15%. Several reports have questioned the validity and reliability of coursework marks. This study sought to investigate this through a quantitative analysis of marks for the nine years included in the study accompanied by interviews with key players in the system: thirteen teachers of physics. The interviews focused on broader aspects of assessment and aimed to analyse the interviewees' thoughts on the reliability, validity and credibility of the school-based assessment (SBA) in light of the changes to be implemented. The results show a weak to moderate positive correlation between the examination mark and the SBA score. Similar results were obtained when comparing the SBA score and the marks scored in practical-oriented questions set in the exam papers. Teachers see the practical aspect in physics as very important but have a number of concerns about its SBA.

Keywords: coursework; school-based assessment; practical work; physics

Introduction

Curriculum innovation is often regarded as an essential part of educational reform. However unless assessment systems change in tandem with the curriculum, the intended curriculum may in practice not be enacted in the classroom and the goals identified not attained. In Malta the National Curriculum Framework (NCF) published in 2012 was seen as a response to the changing demands of society and to reforms in the local education system. A student-centred and inquiry-based pedagogy was indicated as most suited to reflect the rationale of the NCF. But the existing assessment system with high stakes examinations often led to fragmentation of knowledge, rote learning and an increase in teacher-centred pedagogy (Osborne & Dillon, 2008). This assessment culture urgently needed to be addressed. Derived from the NCF, a Learning Outcomes Framework (LOF) was developed which identifies the learning outcomes for all school years and learning areas leading to the development of learning and assessment programmes for the different subjects and levels.

The LOF is attempting to change the assessment culture in schools by including school-based assessment (SBA) as part of the Secondary Education Certificate (SEC) examinations, the national end of compulsory school summative assessment (Attard Tonna & Bugeja, 2016). This will consist of coursework that students carry out during the last three years of compulsory secondary education.

In physics, the contribution of a coursework mark towards the final mark is not a completely new idea. For several years, physics had a coursework component marked by the teachers of the respective students. This consisted of a number of experiments carried out over the last three years of compulsory school, carrying 15% of the marks of the examination.

In the proposed LOF, candidates will be assessed through the coursework and one examination paper. For physics, the examination paper will carry 70% of the marks of the examination while the SBA will carry the remaining 30%. The change is intended as a way of introducing more valid and engaging methods of assessment which have often been criticised as missing in school science across Europe (Osborne & Dillon, 2008). This implies an increase in the weighting of the coursework component, giving the teacher a greater influence

on the students' final mark. Fairness and authenticity are very important considerations in these circumstances (Yates & Johnston, 2018).

The study reported in this paper sought to investigate issues related to assessment, fairness and authenticity as perceived by those who, in the long run, will implement the new assessment system.

Background

Assessment is a fundamental component of the teaching and learning process (Gipps, 1994; Broadfoot & Black, 2004; McColskey, O'Sullivan & Butler, 2005; Gustafsson & Erickson, 2013; Chetcuti & Cutajar, 2014; Black, 2013, 2015; Reiss, 2018) due to its potential contribution towards the improvement of learning (Black & Wiliam, 1998, 2003). However, assessment is also considered as a “contentious feature of education” (Black, 2014, p. 487).

Educational assessment involves the collection of evidence for interpretation and, in turn, communication to the intended users (Harlen, 2005b). Harlen (2009) identifies four essential qualities of any assessment: validity according to its purpose, reliability, positive effect on learning, and cost efficiency (teaching and time). Buhagiar (2007) focuses on trustworthiness and proposes four key assets for a trustworthy assessment that are “credibility, transferability, dependability and authenticity” (p. 44). The complete fulfilment of all the above qualities sounds like a ‘utopia’ for any assessment practice as these qualities are often in conflict and the increase in one quality might hinder the other (Harlen, 2009).

Assessment, in general, can serve two main functions, assessment of learning (summative) or assessment for learning (AfL) (formative). From their earliest introduction it was clear that the terms ‘formative’ and ‘summative’ applied not to the assessments themselves, but to the purposes they serve (Black & Wiliam, 2003). Assessment can only achieve its function when it is carefully designed to fit a clear purpose (Manitoba Education, Citizenship and Youth, 2006). This sometimes gives rise to tensions between the different purposes of assessment (Broadfoot & Black, 2004).

Summative assessment (SA), through the use of tests and examinations, has for many years been considered as the fairest way of ranking students according to their performance in these assessments. However negative effects of examinations have also often been acknowledged such as “teaching to test”

leading to shallow learning (Gipps, 1994, Chetcuti & Grima, 2001; Miller, Linn & Gronlund, 2013). Through the combination of assessment of learning with AfL, SBA is being promoted worldwide (MECY, 2006; Cheng, Andrews & Yu, 2010, Black, 2013) as a way of reducing the adverse effects of examinations (Grima, 2002; Broadfoot & Black, 2004; Harlen, 2004; Murchan, 2018).

The terms SBA and coursework have been used interchangeably in recent reforms in Malta. Coursework is defined as assessment tasks given at regular periods during the course of study (Cambridge University Press, 2020) while SBA involves tasks marked by teachers in the school, with the marks added to other results obtained in external exams (UCLES, 2017). In Malta, coursework is school-based and is assessed by teachers.

SBA allows for a greater range of tasks to be used which increases the validity of such assessment methods (Grima & Ventura, 2000; Harlen, 2004, 2005b, 2009; Wyatt-Smith, Klenowski & Gunn, 2010; Cheng et al., 2010; Johnson, 2013; Yates & Johnston, 2018) compared to examinations which focus on a narrow range of demands (Gipps, 1994; Harlen, 2005b; Black, 2014). When coursework is designed with a purpose, it is likely to enhance student learning, increase motivation to learn and affect teaching strategies (MECY, 2006). SBA gives teachers an enhanced professional role (Harlen, 2005a; Johnson, 2013), but it also means that teachers need to develop skills that will help them gather information for summative purposes while also using it for student learning (Harlen, 2005a). This places teachers at the centre of conflicting purposes (Yates & Johnston, 2018) due to their dual role of providing guidance and support while at the same time giving a judgement (Parker & Volante, 2009).

SBA also raises issues related to the quality of assessment, such as reliability, reference points, validity and record-keeping (MECY, 2006). Teachers are usually confident in their ability to assess accurately although researchers are often concerned about whether teachers can judge student achievement objectively and accurately (Martínez, Stecher & Borko, 2009) and whether such assessments truly reflect fair and authentic learning aims and practices (Black, 2000; Harlen, 2004, 2009; Miller et al., 2013). Reliability is affected by weak application of standards, systematic bias and poorly designed tasks (Morgan & Watson, 2002). Another possible threat is the relationship between teacher and students (Harlen, 2004; Parker & Volante, 2009; Miller et al., 2013; Chetcuti & Buhagiar, 2014) and instances where teachers dismiss the marking scheme in favour of 'thinking' of the students (Wyatt-Smith et al., 2010, p. 71).

Coursework makes it possible to assess skills that cannot be easily assessed during an examination, such as higher order thinking skills, hence improving the validity of the assessment. However, reliability is likely to be negatively affected since such questions or tasks are not easily assessed (Harlen, 2005a). Consequently, SBA is viewed as having high validity but questionable reliability (Harlen 2005a; Wyatt-Smith et al., 2010; Johnson 2013; Murchan, 2018). This, in turn, supports the argument that increasing the validity is likely to result in a drop in reliability (Harlen, 2004, 2005a, 2009; Black, 2000, 2014). The importance of reliability and validity depends solely on the purpose of assessment. Mastering both is difficult (Harlen, 2005a, 2009). This led many countries who have for many years based their assessment on SBA to question the validity of this practice (Johnson, 2013).

Different countries have their own specific SBA policies and practices which impact teaching and learning (Black & Wiliam, 2005). These range from a policy of having no SBA at all in the US (Black & Wiliam, 2005) to 100% SBA in some Australian states (Darling-Hammond & McCloskey, 2008). Other countries show different ways and different extents of the use of SBA and examinations for assessment and certification. The extent of 'trust' seems to vary too. The context including the school system and culture are likely to play an important part in the implementation. Assessment is clearly a complex matter, with many intricate interactions. It is very difficult to capture all of these interactions and produce the most fitting assessment.

In view of the planned reforms, well-prepared strategies are needed to ensure that teaching and learning across the education system are improved. Research carried out in Malta in the past has highlighted discrepancies between the marks obtained in the SBA and the written examination in various subjects, prompting concerns about possible issues of reliability (Grima & Ventura, 2000; Musumeci & Farrugia, 2004; Briffa, Farrugia, & Musumeci, 2005; Grima, Camilleri, Chircop & Ventura, 2005). The study reported in this paper sought to explore these issues further. Thus, the research questions investigated in this study were:

1. (i) What is the correlation between students' school-based assessment scores and the marks obtained in the written examination of the physics SEC?

- (ii) What is the correlation between the school-based assessment mark and students' score in questions related to practical work?
2. What are teachers' attitudes, beliefs and concerns about school-based assessment practices in physics SEC examinations?

The study attempted to tackle the issue of trust with reference to the physics SEC examination by correlating the marks obtained by students in the SBA with marks for other components of the examination. It also attempted to tackle the issue of trust with physics teachers who will be expected to implement the new syllabus and assessment, and obtained their views about the upcoming change and the increased weighting of SBA.

Methodology

A triangulated approach was adopted through the use of multiple methods to enrich the general understanding of the phenomenon studied (Cohen, Manion & Morrison, 2007; Singleton & Straits, 2010; Burton & Obel, 2011). This allows one to "see the same thing from different perspectives and thus to be able to confirm or challenge the findings of one method with those of another" (Laws, Harper & Marcus, 2003, p. 281). While statistics can help to investigate one reality, the qualitative approach gives voice to some of the actors involved in the process. The research design employed helped to compensate the limitations of one method with another (Weick, 1979) such that in combination they enriched what is known about the given research questions (McGrath, 1981).

In order to answer the first research question, statistical tests were used to analyse and correlate the marks obtained by candidates for the SBA and physics written papers for the examination session of nine years. Next, past examination papers were analysed to identify questions that were of a practical orientation. The marks attained in such questions were compared statistically to the scores that the students were awarded for their coursework. The analysis was carried out using Statistical Software Package for Social Sciences (SPSS) 20.0.0. The whole cohort of students sitting for physics during these examination sessions were involved in the statistical analysis. The anonymised marks were provided by the MATSEC examination board, responsible for the SEC certificate in Malta. The board provided results from 2010 to 2019 with the exception of 2011 which was not included in the analysis. For the 2010 session only the total exam marks were available and so the 2010 results could not be used in parts of the analysis.

Following the statistical analysis, semi-structured interviews were carried out with the main actors in this system. This helped in developing more in-depth and flexible data while seeking insight into the experiences, views and beliefs of teachers (Gay, 1976; Osborne & Collins, 2001). A qualitative approach was deemed appropriate to address the second research question, because of the goal to comprehend particular individuals within a particular context (Bryman, 2012, 2016).

An interview guide for a semi-structured interview was prepared. This provided some structure to the interview but at the same time gave the respondents freedom in formulating their answers. The interview questions focused on broader aspects of assessment and probed teachers' thoughts on the reliability, validity and credibility of the SBA based on their experience with the current system and in view of the change that will be implemented in the future. A total of 13 interviews were held, representing males and females with teaching experience ranging from four to thirty-three years. The 13 interviewees were selected from the teachers who volunteered to take part in the research. Two were selected from independent, five from church and six from state schools to reflect the current student population in each sector. The interviews were around thirty minutes long.

After completion of the interviews, transcripts were prepared from the audio recordings. These were coded and the emerging themes were classified into three broad themes: teachers' views about the current coursework system; reliability issues and views about the SBA changes through the LOF. Each broad theme was related to an area tackled by the research questions.

Results

The quantitative analysis

The data used in the quantitative analysis involved the whole population of students registered for the SEC physics examination for each year. Data related to absent students were manually removed so as not to interfere with the statistical analysis. Since the sample used was large when compared to the population, a 99% confidence level was selected. The confidence interval for each year was calculated. This ranged from 0.34 to 0.56.

Over the span of nine years the mean mark obtained in the written examination was 51.10% whereas that obtained in the coursework was 83.08%. The mean total mark was 55.90%. While the minimum and maximum marks and the mode for the written papers varied from year to year, that cannot be said for the coursework component. The modal mark for the coursework component was always 14 out of 15 (93%).

Correlation between students' school-based assessment scores and the marks obtained in the written examination of the physics SEC

This analysis investigated the relationship of two continuous variables, the total written examination mark and the SBA. The Spearman correlation test (non-parametric) was used to investigate the relationship between these two variables since the examination mark distribution is skewed and not normal.

The Spearman correlation coefficient values obtained vary between 0.418 and 0.595, indicating a weak to moderate positive correlation (See Table I). The hypothetical frequencies, (p-values), in all instances are approximately equal to 0, which is less than the 0.05 level of significance indicating a significant relationship between SBA and written examination marks. Therefore, this positive relationship is not attributed to chance and one may expect that the higher the mark in the written examination, the better the candidates will perform in their coursework component.

The pre-LOF physics syllabi and the new LOF syllabus prioritise practical work to help students develop experimental skills. These include the planning and carrying out of investigations, safe and accurate practical techniques, accurate recording and interpretation of data, communicating the data in a clear and accurate manner, and drawing conclusions from the data (MATSEC, 2021a, 2021b). Some of these skills are also essential when answering certain questions that feature in the written part of the physics SEC examination. Examination papers from 2012 to 2019 were analysed for any questions that fit some of these skills. Nine Paper 1 questions were selected and included in this analysis.

The mean coursework mark for these years was always above 81%, while the mean mark of the practical-related questions varies from 33% to 68%. The mode also varies significantly for each exam practical-related question although the coursework modal mark was always 93.33%.

The Spearman correlation test (non-parametric) was used to investigate the relationship between the two continuous variables as the SBA mark distribution is not normal.

Table I: Relationship between marks for written examination and coursework score by Spearman's correlation coefficient

Year		Coursework (15%)	
2010	Paper 1 and 2 combined (85%)	Correlation Coefficient	.418
		p-value	.000*
		Sample size	4143
2012	Paper 1 and 2 combined (85%)	Correlation Coefficient	.500
		p-value	.000*
		Sample size	3780
2013	Paper 1 and 2 combined (85%)	Correlation Coefficient	.521
		p-value	.000*
		Sample size	3664
2014	Paper 1 and 2 combined (85%)	Correlation Coefficient	.489
		p-value	.000*
		Sample size	3599
2015	Paper 1 and 2 combined (85%)	Correlation Coefficient	.496
		p-value	.000*
		Sample size	3285
2016	Paper 1 and 2 combined (85%)	Correlation Coefficient	.545
		p-value	.000*
		Sample size	3343
2017	Paper 1 and 2 combined (85%)	Correlation Coefficient	.595
		p-value	.000*
		Sample size	3003
2018	Paper 1 and 2 combined (85%)	Correlation Coefficient	.581
		p-value	.000*
		Sample size	2902
2019	Paper 1 and 2 combined (85%)	Correlation Coefficient	.587
		p-value	.000*
		Sample size	3039

* $p < .05$

Correlation between the school-based assessment mark and students' score in questions related to practical work.

The results are presented in Table II. The correlation coefficient values range between 0.377 and 0.509 showing a weak to moderate positive correlation. The p-values are less than the 0.05 level of significance indicating that this positive

relationship is significant and not attributed to chance. Hence, one can expect that the higher the mark in the coursework component the better the candidate will perform in practical-related questions found in the written examination although the correlation coefficients are not very strong.

Table II: Relationship between coursework and practical-related questions by Spearman's correlation coefficient

		Coursework		
		Sample Size	Correlation Coefficient	p-value
2012	P1 No.5	3780	.377	.000*
2013	P1 No.7	3664	.453	.000*
2014	P1 No.5	3599	.433	.000*
2015	P1 No.5	3285	.455	.000*
2016	P1 No.5	3343	.455	.000*
2017	P1 No.5	3003	.500	.000*
2018	P1 No.5	2902	.484	.000*
2019	P1 No.2	3039	.509	.000*

* $p < .05$

The Quantitative Analysis

The results from the statistical analysis were shared with the teachers participating in the interviews. They were asked to comment on the strengths and limitations of the current coursework system.

Teachers' attitudes, beliefs and concerns about school-based assessment practices in physics SEC examinations.

Teachers see the coursework component as beneficial as it “not only motivates the students but provides a hands-on approach and puts the theory into practice”. Including coursework as part of the physics SEC encouraged teachers to give greater importance to practical work. Teachers see practical work as a defining characteristic of science, as suggested by Reiss (2018). However, they also mentioned other benefits such as the way the coursework component helped students learn new skills, while decreasing the stress caused by high-stakes examinations since, in their words, “students’ performance is measured throughout the three-year period, not just on how the student

performs on a single day” and “coursework can be done at the students’ pace, so they are more aware of what they are doing”.

Similarly, teachers noted that practical work provides students with an increased sense of ownership and autonomy. Teachers remarked that practical work can help students acquire leadership skills while working in groups.

When considering the proposed reforms, most participants agreed with the doubling of the weighting of the SBA. However, they argued that changes are needed so that the new system will function better than the current one. Maltese teachers identified several limitations in the current coursework system, ranging from the lack of equipment available in schools, inflation of marks, teachers’ input, and the distribution of marks found in the sample marking scheme.

Lack of equipment in schools limits the overall experience of practical work. Teachers often repeat the same experiments each year while sometimes demonstrations replace hands-on student experiments. Repetition, copying and a rushed experience were reported. These have also been observed in other systems with high-stakes assessments (Osborne & Collins, 2001).

The fact that practical work is part of the coursework for SEC physics may also be a disadvantage as teachers and students may focus primarily on scoring high marks. This could be the reason why secondary school science “has concentrated far too much on the controlled and readily reproducible experiments” (Jenkins, 2002, p.158). Indeed, this might also lead to a recipe-following format where students simply follow instructions without showing any understanding about what they are doing during the practical sessions (Reiss, 2018). In the interviews, teachers’ input during the practical session was also mentioned:

I am sure that no teacher lets the students perform experiments without giving them input and guidance throughout”.

“Most teachers, including myself, we would want our students to do well in their coursework since it makes up part of their MATSEC exam.

This input may be of various forms, such as dictating reports, and was given as a reason for the apparent inflation of marks detected in the quantitative analysis.

Another issue that emerged is the role of teachers during the practical session. Teachers highlighted a dilemma encountered since teachers are expected to help their students improve but at the same time they are expected to stand back and assess during the practical sessions:

I think there is a dilemma because you want a formative assessment, you do not want it to be a punishment ... with formative assessment you try to encourage possibly ...

Teachers mentioned the lack of uniformity, where assessment practices vary amongst different teachers. One of the teacher remarked that “some students are at an advantage because their teachers give marks without going through the actual criteria, whilst other teachers make sure that their assessments are accurate and fair”. Another teacher said “I have seen reports of students from other classes that do not even deserve the mark of 9 but they still obtain 14 ... not everyone ensures that the marks obtained reflect the student’s real performance”.

This lack of uniformity does not appear to be limited to the Maltese context. When referring to such issues, Cumming and Maxwell (2004) argue that these problems are attributed to a lack of common understanding about the assessment being carried out. Therefore, it is imperative that teachers are given clear guidelines and that they follow the guidelines especially through the marking scheme. Teachers criticised the marking criteria suggested in the syllabus which focus mainly on the format of the experiment report rather than on physics or on practical skills: “the marks are not all focused on physics material but rather on the format of the report and diagram drawing. Then the precautions and conclusion ...”. This could be one reason for the lack of uniformity mentioned in the interviews.

Participants appeared confident in their own ability to assess coursework accurately however they were more likely to be concerned at the way that other teachers assess. This is not a unique situation. Research suggests that although teachers are usually confident in their ability to assess correctly (Martínez, Stecher & Borko, 2009) there are concerns as to whether teachers’ judgement can truly reflect fair and authentic practices (Black, 2000; Harlen, 2004, 2009; Miller et al., 2013). Participants suggested the introduction of moderation between teachers as internal verification within the schools, similar to the approach implemented for SEC vocational subjects in Malta. This would be a form of inter-rater reliability exercise (Johnson, 2013).

Discussion

While the statistical analysis shows a positive correlation between SBA and written examination marks, one can note that the coursework marks seem to be inflated and with a low variance between candidates. This has been suggested in other local research studies (See: Grima & Ventura, 2000; Musumeci & Farrugia, 2004; Briffa et al., 2005; Grima et al., 2005; Pirota & Musumeci, 2012). One may argue that the correlation is not expected to be too high since the two components assess different skills. After all the SBA component in physics was introduced to assess skills that cannot be examined through traditional exams (Grima et al., 2005).

Participants were in favour of including practical work as part of the coursework for SEC physics and gave several reasons such as increased motivation, improved attitudes towards the subject and the acquisition of several skills. Toplis (2012) reported that in a study carried out in England, students found practical work important, giving three main reasons: it instils interest through participation and autonomy; provides an alternative form of teaching; and as a way of learning to facilitate recall and memorisation. While, undoubtedly, the practical tasks carried out for the SEC examinations might contribute to many benefits, this is highly dependent on the kind of practical work students experience. A variety of experiments that require a diversity of skills is clearly needed (Musumeci and Farrugia, 2004).

The limitations put forward by the teachers, such as difficulties with assessment of practical work, confirm that practical work is not easily evaluated due to its complexity (Abrahams & Reiss, 2015). The current marking criteria focus mainly on the indirect assessment of practical skills (IAPS) which emphasise what students know about practical work and how it should be done rather than on actual practical skills (Abrahams & Reiss, 2015). Including direct assessment of practical skills (DAPS) would enhance the validity of the SBA. This limitation may be linked to the relationship between the SBA mark and practical-oriented questions in the written examination. The Spearman correlation test showed a weak to moderate positive correlation between the SBA mark and students' scores in questions related to practical work. This positive correlation might indicate that the coursework component is helping students in acquiring skills that can be assessed by IAPS which are reproducible in the written examination component, as suggested by the participating teachers. However, the practical component also involves skills that need to be assessed by DAPS (Abrahams & Reiss, 2015). In a local study,

Pirotta and Musumeci (2012) observed that through the practical sessions, students acquired skills which feature in the written examination (such as graph-plotting) while they lacked practical skills that should develop through coursework. They concluded that the coursework does not always result in the intended learning outcomes which raises concerns about the validity of such practices (Pirotta & Musumeci, 2012). This observation could be related to the use of teacher demonstrations to replace student hands-on experiments, mentioned by teachers, which limits the development of certain practical skills. Another possible reason is the high-stakes nature of the assessment which may lead to a shift towards “teaching to test” emphasising skills that are examinable during the written examination.

The need for clear assessment criteria and for training teachers in the use of these criteria to increase reliability (Grima & Ventura, 2000; Grima et al., 2005) still exists. A change in the current moderation process from a one-off end of course judgement to a continuous learning dialogue between moderators and teachers was also suggested (Ventura & Murphy, 1998; Grima & Ventura, 2000; Grima et al., 2005).

Another interesting point that emerged is related to teachers’ dual role as teachers and examiners. Under such circumstances teachers may need help with finding a balance between FA and the assessment for certification purposes.

Some of the internal weaknesses in the system appear to have been tackled in the documents available for the new physics LOF (MATSEC, 2021b). These address both validity and reliability issues. The introduction of new assessment tasks encourages development of a wider set of skills through practical work. Marking criteria have been updated and teachers are required to present a breakdown of marks for each task presented. The marking criteria should also incorporate a fine breakdown of marks so as to minimise differences in interpretation. Moreover, the criteria should assess both IAPS and DAPS.

Teachers seem to lack trust in the way that other teachers assess coursework. Similar views were reported elsewhere, and several reasons have been given for concerns as to whether teachers’ judgement can truly reflect fair and authentic practices (Black, 2000; Harlen, 2004, 2009; Miller et al., 2013). For example, teachers may be influenced by students’ behaviour, socio-economic

background, and effort (Morgan & Watson, 2002; Harlen, 2004, 2005a; Briffa et al., 2005; Martínez et al., 2009; Ready & Wright, 2011; Brookhart, 2013; Johnson, 2013) or even the relationship between teacher and students (Harlen, 2004; Parker & Volante, 2009; Miller et al., 2013; Chetcuti & Buhagiar, 2014). Although moderation already takes place in schools, yet the issue of reliability persists (Grima & Ventura, 2000; Musumeci & Farrugia, 2004; Grima et al., 2005). It is envisaged that with the new LOF, moderation will take place every year during the three-year physics course rather than at the end of the course (MATSEC, 2021b). For a system to be strong and fulfil its intended outcomes, support and properly resourced moderation practices are required (Cumming & Maxwell, 2004). While the proposed increase in frequency of moderation is an improvement, there needs to be a shift in the type of moderation being held. Moderation should lead to a constructive dialogue between moderators and teachers to support teacher development, as suggested by Harlen (1994). This is even more important if moderation is carried out throughout the course.

The moderation process serves as a comparison between schools, conducted by moderators sent by the examination board through a “human benchmark exercise” (Johnson, 2013, p. 94). Some teachers in the interviews suggested internal verification within schools where teachers are paired together and verify each other’s work. These changes might prompt growth in collegiality amongst the teachers which in turn instils a higher sense of trust in the SBA system that will be implemented. This seems to be essential since participating teachers heavily criticised the way that other teachers assess or help their students since this jeopardises the fairness of the assessment.

From the interview data, it seems that since each class has one physics teacher, the teacher feels fully responsible for his/her class and does not know what is happening in other classes. This may result in a lack of trust in the SBA system and in the way that other colleagues are marking. Teacher training is essential in ensuring that valid and trustworthy assessment tasks are presented (Grima et al., 2005; Harlen, 2009; Black, 2015). It is also important that exemplars are used to familiarise teachers with the variety of scenarios that they might run into. This might help to increase reliability and confidence in the assessment system. With the help of the organisers, teachers might be asked to assess a sample of coursework so that as a group the common difficulties in marking are drawn out.

Conclusion

The planned assessment reforms have been described as attempts to change the assessment culture in Maltese schools (Attard Tonna & Bugeja, 2016) which for years has been examination-oriented. The increase in the weighting of the SBA in SEC physics is a welcome change as it highlights the importance of development of skills that go beyond what may be examined in a written examination paper. The new system proposes changes that are likely to influence validity and reliability. It is evident that a number of issues need to be addressed before the coursework component becomes more prominent. This study highlighted a number of systemic weaknesses that need to be addressed. However, while changes are important in a dynamic educational environment, teachers' involvement in reforms is critical for successful implementation (Fernandez, Ritchie & Barker, 2007; Ibrahim, Al-Kaabi & El-Zaatari, 2013) as they are key players in the proposed change. The role of teachers as assessors will increase considerably, thus teachers need to be prepared or else the implementations are likely to fall short (Harlen, 2005a). Any reform will only be successfully implemented if all involved work together for a common objective. On the other hand, resistance is a major factor in reform failure (Zimmerman, 2006) stressing the importance of valuing teachers' beliefs and experiences and ensuring that teachers are well-prepared for any change within the educational system.

References

- Abrahams, I., & Reiss, M. J. (2015). The assessment of practical skills. *The School science review*, 96(357), 40 – 44.
- Attard Tonna, M. & Bugeja, G. (2016). A reflection on the Learning Outcomes Framework Project. *Malta review of Educational Research*, 10(1), 169-176.
- Black, P. (2000). Research and the Development of Educational Assessment. *Oxford Review of Education*, 26(3-4), 407 – 419. DOI: 10.1080/713688540
- Black, P. (2013). Pedagogy in Theory and in Practice: Formative and Summative Assessments in Classrooms and in Systems. In D. Corrigan, R. Gunstone & A. Jones (Eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy* (pp. 207 – 229). <https://doi.org/10.1007/978-94-007-6668-6>
- Black, P. (2014). Assessment and the aims of the curriculum: An explorer's journey. *Prospects*, 44(4), 487 – 501. DOI: 10.1007/s11125-014-9329-7
- Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, (22)1, 161 – 177. DOI: 10.1080/0969594X.2014.999643

- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning, Assessment in Education. *Principles, Policy & Practice*, 5(1), 7-74. DOI: 10.1080/0969595980050102
- Black, P., & Wiliam, D. (2003). 'In Praise of Educational Research': formative assessment. *British Educational Research Journal*, 29(5), 623 – 637. DOI: 10.1080/0141192032000133721
- Briffa, C., Farrugia, J. & Musumeci, M. (2005). *Assessing School-Based material and Non-Written Work*. Paper presented in 6th Association of Educational Assessment (Europe) Annual Conference "Assessment and Accountability". Dublin: Republic of Ireland.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in Education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7 – 26. DOI: 10.1080/0969594042000208976
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69 – 90. DOI: 10.1080/0969594X.2012.703170
- Bryman, A. (2012). *Social Research Methods*. New York: OUP Oxford.
- Bryman, A. (2016). *Social Research Methods*. Oxford University Press.
- Buhagiar, M. A. (2007). Classroom assessment within the alternative assessment paradigm: revisiting the territory. *The Curriculum Journal*, 18(1), 39 – 56. DOI: 10.1080/09585170701292174
- Burton, R. M., & Obel, B. (2011). Computational Modeling for What-Is, What-Might-Be, and What-Should-Be Studies—And Triangulation. *Organization Science*, 22(5), 1195–1202. <https://doi.org/10.1287/orsc.1100.0635>
- Cambridge University Press. (2020). *Meaning of coursework in English*. Retrieved from <https://dictionary.cambridge.org/dictionary/english/coursework>
- Cheng, L., Andrews, S., & Yu, Y. (2010). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221 – 249.
- Chetcuti, D. & Grima, G. (2001). *Portfolio Assessment*. Malta: Ministry of Education.
- Chetcuti, D., & Buhagiar, M. A. (2014). Assessing the field placement in initial teacher education: finding the balance between formative and summative assessment. *Problems of Education in the 21st Century*, 58, 39 - 52.
- Chetcuti, D., & Cutajar, C. (2014). Implementing Peer Assessment in a Post-Secondary (16–18) Physics Classroom. *International Journal of Science Education*, 36(18), 3101 – 3124. DOI: 10.1080/09500693.2014.953621
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education, Sixth Edition*. Taylor & Francis e-Library
- Cumming, J., & Maxwell, G. (2004). Profiles of educational assessment systems worldwide. *Assessment in Education: Principles, Policy & Practice*, 11(1), 89 - 108. DOI: 10.1080/0969594042000209010

- Darling-Hammond, L., & McCloskey, L. (2008). Assessment for Learning around the World: What would it mean to be internationally competitive? *Phi Delta Kappan*, 90(4), 263 - 272. <https://doi.org/10.1177/003172170809000407>
- Fernandez, T., Ritchie, G., & Barker, M. (2007). A sociocultural analysis of mandated curriculum change: the implementation of a new senior physics curriculum in New Zealand schools. *Journal of Curriculum Studies*, 40(2), 187-213.
- Gay, L. R. (1976). *Educational research competences for analysis and application*. Ohio: Charles E. Merrill Publishing Company.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London, UK: Taylor & Francis e-Library.
- Grima, G. & Ventura, F. (2000, September). *School-based assessment in Malta: Lessons from the past, directions for the future*. Paper presented at The First ACEAB Conference, Mauritius.
- Grima, G. (2002). Assessment Issues in Maltese Secondary Schools. In Bezzina, C., Camilleri-Grima, A., Purchase, D. & Sultana, R. (2002), *Inside Secondary Schools: A Maltese Reader* (pp. 137 - 154). Msida: Indigo Books.
- Grima, G., Camilleri, R., Chircop, S., & Ventura, F. (2005). *MATSEC: Strengthening a National Examination System*. Retrieved from https://education.gov.mt/en/resources/Documents/Policy%20Documents/matsec_review.pdf
- Gustafsson, J., & Erickson, G. (2013). To trust or not to trust? – teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, 25(1), 69-87.
- Harlen, W. (1994). Developing Public Understanding of Education: A Role for Educational Researchers. *British Educational Research Journal*, 20(1), 3 - 16.
- Harlen, W. (2004). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. (2005a). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245 - 270. DOI: 10.1080/02671520500193744
- Harlen, W. (2005b). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16(2), 207 - 223.
- Harlen, W. (2009). Improving assessment of learning and for learning. *Education 3-13*, 37(3), 247 - 257. DOI: 10.1080/03004270802442334
- Ibrahim, A. S., Al-Kaabi, A., & El-Zaatari, W. (2013). Teacher resistance to educational change in the United Arab Emirates. *International Journal of Research Studies in Education*, 2(3), 25-36.

- Jenkins, E. (2002). A New Science Curriculum for All? *Studies in Science Education*, 38(1), 155-160.
- Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education*, 28(1), 91-105. DOI: 10.1080/02671522.2012.754229
- Laws, S., Harper, C. and Marcus, R. (2003). *Research for development*. London: SAGE.
- Manitoba Education, Citizenship and Youth (MECY). (2006). *Rethinking Classroom Assessment with Purpose in Mind: assessment for learning, assessment as learning, assessment of learning*. Retrieved from https://www.edu.gov.mb.ca/k12/assess/wncp/full_doc.pdf
- Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom Assessment Practices, Teacher Judgments, and Student Achievement in Mathematics: Evidence from the ECLS. *Educational Assessment*, 14(2), 78 - 102. <https://doi.org/10.1080/10627190903039429>
- MATSEC. (2021a). *SEC 24 Syllabus: Physics 2024*. Retrieved from <https://www.um.edu.mt/matsec/syllabi/secsyllabi2024>
- MATSEC. (2021b). *SEC 24 Syllabus: Physics 2025*. Retrieved from <https://www.um.edu.mt/matsec/syllabi/secsyllabi2023>
- McColskey, W., O'Sullivan, R. & Butler, S. (2005). *How to Assess Student Performance in Science: Going Beyond Multiple-Choice Tests*. Greensboro, NC: Regional Educational Laboratory Southeast.
- McGrath, J. E. (1981). Dilemmatics- The Study of Research of Research Choices and Dilemmas. *American Behavioral Scientist*, 25(2), 179-210. Retrieved from <https://www.cebma.org/wp-content/uploads/McGrath-1981-ABS-Dilemmatics.pdf>
- Miller, D., Linn, R. L., & Gronlund, N. (2013). *Measuring and Assessment in Teaching* (11th ed.). Singapore: Pearson Education South Asia Pte Ltd.
- Morgan, C. & Watson, A. (2002). The Interpretative Nature of Teachers' Assessment of Students' Mathematics: Issues for Equity. *Journal for Research in Mathematics Education*, 33(2), 78 - 110.
- Murchan, D. (2018). Introducing school-based assessment as part of junior cycle reform in Ireland: a bridge too far? *Educational Assessment, Evaluation and Accountability*, 30(2), 91-131.
- Musumeci, M., & Farrugia, J. (2004, April). *School-Based Assessment within the Secondary Education Certificate (16+) Science Exams in Malta*. Paper presented at the CASTME Conference, Cyprus.
- Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical reflections*. (A Report to the Nuffield Foundation). Retrieved from http://efepereth.wdfiles.com/local--files/science-education/Sci_Ed_in_Europe_Report_Final.pdf
- Osborne, J., & Collins, S. (2001). Pupils' views of the role and value of the science curriculum: A focus-group study. *International Journal of Science Education*, 23(5), 441 - 467. DOI: 10.1080/09500690010006518

- Parker, D. C., & Volante, L. (2009). Responding to the Challenges Posed by Summative Teacher Candidate Evaluation: A collaborative self-study of practicum supervision by faculty. *Studying Teacher Education: A journal of self-study of teacher education practices*, 5(1), 33 - 44. <http://dx.doi.org/10.1080/17425960902830385>
- Pirotta, D. & Musumeci, M. (2012, January). *The 'currency' of coursework in national examinations: SEC physics as a case study*. Paper presented at ASE Annual Conference, University of Liverpool: UK.
- Ready, D.D., & Wright D. L. (2011). Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities: The Role of Child Background and Classroom Context. *American Educational Research Journal*, 48(2), 355 - 360.
- Reiss, M. J. (2018). Beyond 2020: ten questions for science education. *The school science review*, 100(370), 47-52.
- Singleton, R., & Straits, B. C. (2010). *Approaches to Social Research*. Oxford University Press.
- Toplis, R. (2012). Students' Views about Secondary School Science Lessons: The Role of Practical Work. *Research in Science Education*, 42(3), 531 - 549. DOI: 10.1007/s11165-011-9209-6.
- University of Cambridge Local Examinations Syndicate (UCLES). (2017). *Assessment for learning*. Retrieved from <https://www.cambridgeassessment.org.uk/>
- Ventura, F., & Murphy R. (1998). The impact of measures to promote equity in the secondary education certificate examinations in Malta: An evaluation. *Mediterranean Journal of Educational Studies*, 3(1), 47 - 73.
- Weick, K. E. (1979). *The Social Psychology of Organizing*. McGraw-Hill.
- 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. DOI: 10.1080/09695940903565610.
- Yates, A., & Johnston, M. (2018). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 25(6), 638-654. DOI: 10.1080/0969594X.2017.1295020
- Zimmerman, J. (2006). Why some teachers resist change and what principals can do about it. *NASSP Bulletin*, 90(3), 238-249. doi: 10.1177/0192636506291521