

## Relative Difficulty of Subjects at Secondary Education Certificate Level

**Gilbert J Zahra, Dario Pirotta, Frank Ventura**

MATSEC Support Unit, University of Malta

Gilbert.j.zahra@um.edu.mt

### **Abstract**

It is commonly held that an examination body should maintain a standard level of difficulty across different years, tiers, and subjects. Grade setting does depend, to a certain extent, on expert judgement and, not surprisingly, different studies have suggested that the same standard of difficulty is not maintained across different examination boards and subjects. In this study the level of difficulty of subjects is measured by comparing the mean general ability of candidates who obtain the same grades in the different subjects at Secondary Education Certificate (SEC), which are offered by the MATSEC Examinations Board of the University of Malta. The research method and the ensuing results are explained in detail and discussed. The outcomes show that although some differences between subjects are present, with one subject in particular being flagged as being possibly graded much easier than other subjects, differences between most subjects were not significant enough to allow a clear ordering in terms of difficulty.

### **Introduction**

The MATSEC Examinations Board was founded in 1991 with the aim of replacing the GCE examinations at Ordinary and Advanced levels offered by foreign examination boards and rationalising the few subjects offered by the University of Malta at Ordinary and Advanced Matriculation levels. Gradual changes took place over three years and, by 1994, the board replaced the GCE Ordinary level subjects with a set of subjects at Secondary Education Certificate (SEC) level with a different format to the previous Matriculation subjects. In the new format all subjects consist of written examinations in two papers with Paper 1, a common paper taken by all candidates, and an option of either Paper 2A or Paper 2B. Paper 2A consists of more challenging questions than Paper 1

while Paper 2B consists of less challenging questions. Candidates who sit for Paper 1 and Paper 2A could qualify for Grade 1 to 5, with Grade 1 being the highest grade, and those performing lower than Grade 5 remain unclassified (Grade U). Candidates sitting for Paper 1 and Paper 2B could qualify for Grades 4 to 7 or remain unclassified (Grade U). Language subjects have an externally assessed oral and aural component in Paper 1 while science subjects and other subjects of a practical nature have a school assessed coursework component that is externally moderated. Further changes took place by 1997 when the MATSEC Board offered examinations at Advanced and Intermediate Matriculation levels leading to the Matriculation Certificate, which was modelled on the International Baccalaureate diploma qualification. The focus of this paper is on the relative difficulty of SEC level subjects only.

The awarding of grades is based on professional, expert judgement about candidate performance aided by statistical measures. This process is similar to that used by most foreign examination boards which offer certification at the same levels as the MATSEC Examinations Board (Gardner, 2013; Good & Cresswell, 1988) particularly those in England (Newton, 2007). The MATSEC Support Unit, which is the operational arm of the MATSEC Board, has developed and upgraded its quality assurance standards over the years. Moreover, the Support Unit has also outlined its procedures for grade boundary setting and made these available to the general public (MATSEC Support Unit, 2015).

Educational assessment has always been contestable. Gardner (2013) argues that the term measurement is used too loosely in the field, prompting the belief that one is able to measure abstract constructs like logic, skills, competences, and reasoning and describe such complex characteristics using a single number or grade. However, while in science, a measurement is a multiple or a fraction of a standard unit (e.g. the room is four times as wide as the metre, which is a standard unit of length), in educational assessment there is no single agreed unit of 'knowledge' or 'logic'. Indeed, in any measurement there are several potential sources of error, including personal errors and instrumental errors. A statistical estimate of these errors can be calculated and reported as the standard error of measurement. In the case of examinations, the standard error of judgement implies that the grade awarded to a number of candidates could be potentially misjudged. This has been well illustrated for example by research carried out by Gardner (2013) in which experts from different examining bodies were asked to rate papers from other bodies. It is therefore

understandable that the process of educational assessment and grade awarding remains contestable on a number of grounds, either due to media sensationalism (Gardner, 2013) or the inevitable subjectivity of expert judgement (Gardner, 2013; Newton, 2007; Good & Creswell, 1988).

Tiered assessments add complexity to the issue (Newton, 2007; Good & Creswell, 1988). While Grades 1-3 in SEC examinations can only be obtained by candidates who opt for the more challenging Paper 2A option, and Grades 6-7 are only obtainable by candidates who choose the less challenging Paper 2B option, Grades 4-5 (and Grade U) are achievable through both routes. Thus, one is bound to question whether a Grade 4 attained by a candidate who sat for Paper 2A is equivalent to a Grade 4 obtained by a candidate who sat for Paper 2B of the same subject.

Besides the question of comparability of identical grades achievable through different routes within the same subject, questions of inter-subject comparability exist. Although each subject's syllabus content and assessment are designed to address different areas of knowledge and skills, it is common for candidates or the general public to assume that one subject is less challenging than another. Chemistry, for instance, is a subject commonly referred to as a difficult subject (Childs & Sheehan, 2009; Crippen & Brookes, 2008), especially by candidates who are not studying the subject (Barbara, Muscat, & Zahra, 2010). Such observations might be more than mere perceptions and research by Fitz-Gibbon and Vincent (1997) concluded through four different measures that Advanced Level examinations of Mathematics, science subjects, and foreign languages are more 'severely graded' than other subjects. Coe (2008), reviewing the relative difficulty of UK General Certificate for Secondary Education (GCSE) subjects, reached similar conclusions, which are summarized in the figure below.

But how does one define subject difficulty? Fitz-Gibbon and Vincent (1997) argue that subject difficulty is a construct. They refer to a subject as being difficult when the grades awarded to candidates are lower than expected as shown by adequate statistics. Similarly, Coe (2008, p.613) maintains that "rather than saying that Maths is 'harder' than English we must say that a particular grade in Maths indicates a higher level of general academic ability than would the same grade in English."

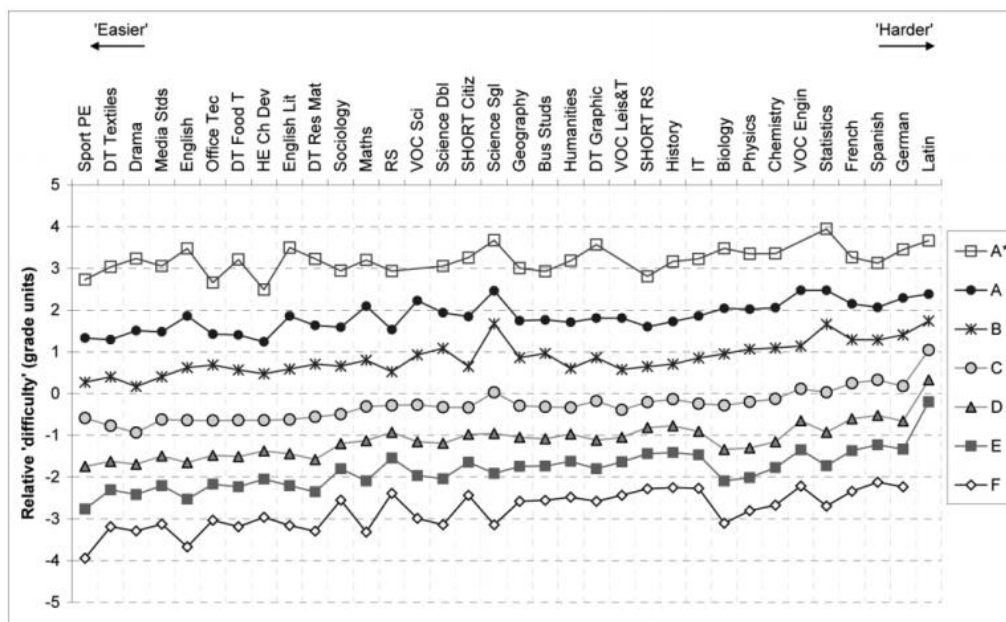


Figure 1: Relative 'difficulties' of achieving grades A\* through F in 34 GCSE subjects, ordered by weighted average difficulty (Coe, 2008, p.625)

As has happened since at least the 1970s, when research shed doubts on inter-subject comparability, criticism has always been directed to these academic endeavours (Newton, 2012). Coe (2008) summarizes six arguments against the use of statistical models for comparing inter-subject difficulty.

1. Candidate performance is affected by subject specific qualities which may include student interest, teaching quality, external motivators, curriculum time, and so on.
2. Different groups of candidates might opt for different groups of subjects. Data by Coe (2008) suggests that this might be the case for subjects which require a performance, such as Physical Education (P.E.) and Music, and atypical languages such as Urdu. This is comparable, for example, to the relatively high performance of candidates sitting for Russian at all levels of MATSEC examinations and which is frequently observable in SEC examinations statistical reports.
3. Differences between sub-groups of students exist and particular groups of students may be over- or under-represented in some subjects rather than others. For instance, at the University of Malta, working class students tend to be over-represented in courses such as education and

engineering and under-represented in courses such as medicine, law and architecture (Sultana, 1995 cited in Cutajar, 1999).

4. Different methods of estimating subject difficulty can provide different results.
5. Forcing subjects to be graded with equal difficulty would be problematic as it would result in higher pass rates in subjects which are harder (or attracting a more able population of candidates) and much lower pass rates in subjects which are easier (or attracting a less able population of candidates).
6. Different subjects are simply not comparable. They cannot be 'harder' or 'easier' but merely different, requiring a different set of skills and competences altogether. For instance, if an examination board offers two different syllabi and assessments for a language, one for native speakers and one for foreign speakers, it cannot be expected that candidates being awarded the same grade from the examinations have comparable skills and competences.

Although a thorough discussion of the arguments above goes beyond the scope of this article, it is important to remember that questions in social research are seldom dichotomous ones. Subjects cannot be assumed to be fully comparable or not comparable (Fitz-Gibbon & Vincent, 1997). As examinations share a similar structure and report the same grades, some inter-subject comparability is inevitable (Newton, 2012; Coe, 2008; Fitz-Gibbon & Vincent, 1997). Besides, as current practices for processing admission requirements for further studies or recruitment for employment equate identical grades obtained in different subjects, the absence of comparability between the grading of different subjects would be unfair (Coe, 2008).

## **Methodology**

This study aims to shed light on the relative difficulty of subjects by considering the general academic ability of candidates obtaining the same grades in different SEC examinations. Thus, this study seeks to ascertain the level of inter-subject comparability by adopting a specific-causes causal definition of the term (Newton, 2012). In brief, such an approach assumes that a single property can be measured and used for comparison purposes to ascertain what it takes for candidates to achieve the same grade in different subjects. Although such an approach is not free from criticism, it also makes most of the criticisms of statistical models for determining inter-subject comparability immediately irrelevant (Coe, 2008).

Given that there is usually a medium to high correlation between candidates' attainment in different subjects, it is possible to speculate that different educational assessments and "different cognitive test approaches appear to [measure] essentially the same construct, namely general national cognitive ability" (Rindermann, 2007, p.687). This assumption is not fully watertight: people can be seen as having different intelligences and interests (Gardner, 2004; Wilson, 1996). For example, Art involves skills which are not demonstrated in subjects where the candidates' expression is limited to writing. As noted earlier, questions in social science do not elicit dichotomous replies (Fitz-Gibbon & Vincent, 1997). Correlation between different educational achievements in different subjects is strong but not perfect. Some truth can be attributed to both explanations of intelligence and seeking which one, if any, is more valid goes beyond the scope of this article. However, this study uses a measure of the candidates' general academic ability which is calculated from their achievement in four subjects considered as core subjects.

All the results of 3147 candidates who sat for SEC examinations in English Language, Maltese, Mathematics and Physics in 2016, besides other subjects, were used for the purposes of this research. An average of the candidates' raw scores in these four subjects, which are subjects with large number of entries, was calculated and assumed to be a measure of their general academic ability, denoted by  $G_4$ . The candidates' attainments in these four subjects show a medium to strong correlation, as expected (Table 1).

Table I: Correlation between achievements in the four subjects used to measure candidates' general academic ability

Correlations	Maths	English	Maltese	Physics
<b>Maths</b>	/	0.68	0.69	0.85
<b>English</b>	0.68	/	0.68	0.72
<b>Maltese</b>	0.69	0.68	/	0.74
<b>Physics</b>	0.85	0.72	0.74	/

A distribution of general attainment across this population of 3147 candidates shows a nearly normal distribution, with a mean score of 57.4 and a skewness of -0.59. This suggests that, for the candidates who registered for at least

English Language, Maltese, Mathematics and Physics, achievement is skewed towards higher than average attainment.

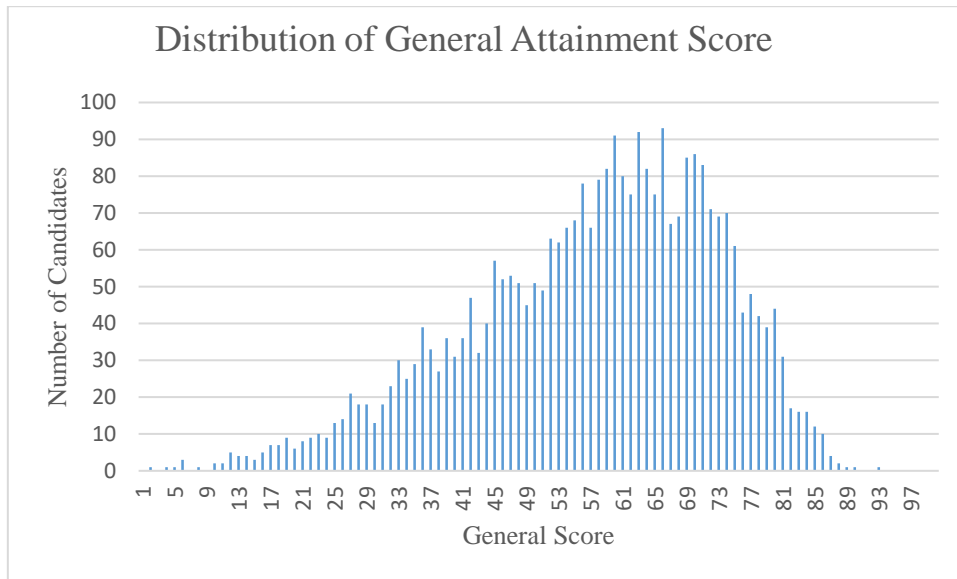


Figure 2: Distribution of Candidates' General Attainment Score

Computations were then carried out of the average G4 scores of candidates who obtained each of the possible grades (Grades 1 to 7 and U) for SEC subjects with 200 candidates or more. These scores are presented in Table II which shows, for example, that in Accounts (Table II) there were 427 candidates and the candidates who obtained a Grade 1 had an average G4 score of 78.3, those who obtained a Grade 2 had an average G4 score of 74.6, and so on until Grade 7 and U. Table III presents the total number of candidates in each subject (in the first column) followed by the number of candidates in each grade. These are the candidates whose G4 scores were used to compute the values in Table II.



Table II: Candidates' average general attainment ( $G_4$ ) in different SEC subjects, by Grade

Subject	N	Grade							
		1	2	3	4	5	6	7	U
Accounts	427	78.3	74.6	70.7	66.2	61.9	57.3	59.7	54.7
Art	447	64.2	65.4	61.5	54.3	51.2	43.4	37.7	40.3
Biology	971	79.9	75.3	70.4	64.7	59.5	53.8	53.2	48.0
Business Studies	288	77.5	72.0	67.6	62.9	56.2	49.1	43.8	41.7
Chemistry	715	81.0	75.7	71.2	67.4	62.3	60.0	57.4	55.1
Computer Studies	704	78.4	71.4	65.8	57.9	51.8	42.9	39.3	39.6
Design & Technology	293	72.1	66.2	58.2	57.6	50.9	36.3	33.4	44.4
English Literature	2049	79.1	75.1	70.2	64.9	59.9	54.7	51.4	50.3
Environmental Studies	1182	79.4	75.9	70.2	65.6	59.5	52.3	46.2	48.4
French	992	78.2	73.9	68.0	64.6	59.5	54.8	52.6	49.6
German	343	78.8	74.6	65.4	61.6	57.2	48.9	42.1	47.8
Graphical Communication	458	74.1	70.8	65.4	62.0	54.5	47.0	45.4	44.7
Home Economics	573	72.9	64.6	55.5	50.9	45.5	36.5	28.2	36.3
Italian	1417	72.5	69.4	63.6	60.1	57.3	50.5	47.4	45.0
Physical Education	282	71.7	66.6	60.3	54.8	50.7	31.5	27.2	37.2
Religious Knowledge	2687	78.9	72.2	66.8	60.8	56.9	48.5	42.4	42.0
Social Studies	650	78.5	74.2	69.6	65.3	59.5	56.3	46.5	48.9
Spanish	249	76.8	71.4	64.9	59.7	55.7	52.6	52.4	46.2



Table III: Number of candidates for whom a general attainment was computable, by Subject and Grade

Subject	N	Grade							
		1	2	3	4	5	6	7	U
Accounts	427	37	59	89	73	61	11	10	87
Art	447	16	42	86	83	99	22	14	85
Biology	971	90	117	183	205	132	39	22	183
Business Studies	298	12	27	38	40	69	20	23	59
Chemistry	715	85	133	135	121	91	12	28	110
Computing	704	49	117	163	150	101	41	23	60
Design & Technology	293	10	22	32	64	52	40	9	64
English Literature	2049	102	172	447	528	296	99	68	336
Environmental Studies	1182	45	107	240	213	242	73	24	238
French	992	100	179	214	154	143	37	41	124
German	343	25	61	87	67	46	21	13	23
Graphical Communication	458	32	55	79	95	105	12	15	65
Home Economics	573	22	100	130	119	42	59	17	84
Italian	1417	91	166	261	292	176	138	104	189
Physical Education	282	15	37	51	59	50	12	3	55
Religious Knowledge	2687	123	324	598	600	368	220	130	324
Social Studies	650	20	49	80	133	144	35	21	168
Spanish	249	6	23	43	51	43	9	19	55

## Analysis and Discussion

Figure 3 below shows the trend in candidate mean  $G_4$  scores across the studied subjects for each grade.

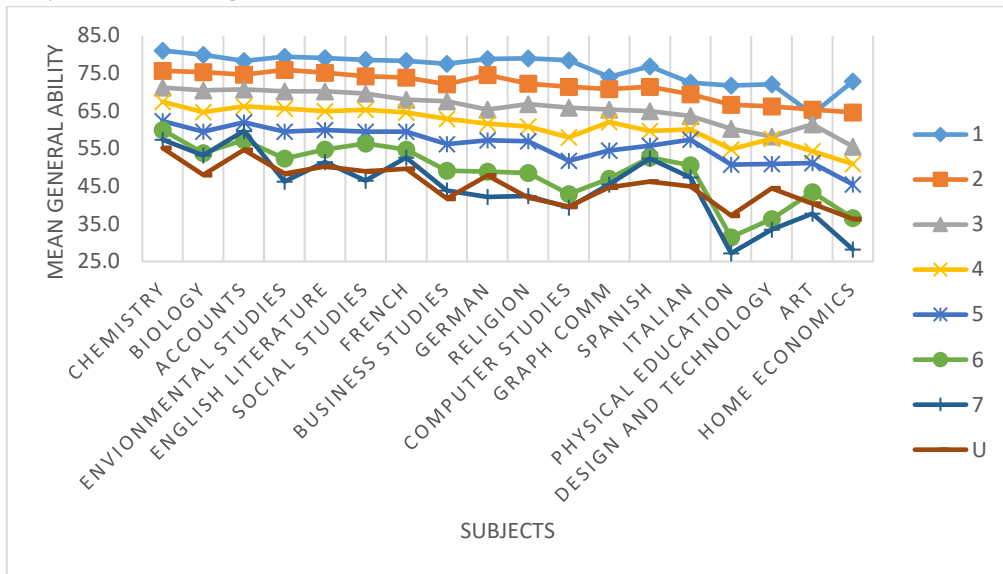


Figure 3: Mean General Attainment Scores across different subjects, per grade

In general, the calculated candidates'  $G_4$  score decreases with the attainment of lower grades irrespective of the SEC subject under consideration. This result supports the idea that there is a general ability which affects candidate attainment in all subjects.

Some exceptions can be noted. Most of them are minor and noted in the lower grades where the average of candidates obtaining an unclassified grade (U) is higher than that of candidates obtaining a better grade. A similar observation was noted by Coe (2008). This discrepancy was noted in Art, Design and Technology, Computing, Environmental Studies, German, Home Economics, P.E. and Social Studies. This is not surprising since one must bear in mind that while all candidates obtaining Grades 6 and 7 sat for Paper 2B, candidates obtaining Grade U could have sat for either Paper 2A or Paper 2B. As candidates who opt for Paper 2A obtain Grade U if their performance in the subject falls below that expected for Grade 5, these candidates could have contributed to the higher than expected  $G_4$  scores for the Grade U candidates.

Two other exceptions can be observed in Accounts, where the average  $G_4$  score obtained by Grade 7 candidates is higher than that obtained by Grade 6 candidates, and Art, where the average  $G_4$  score obtained by Grade 2 candidates is higher than that obtained by Grade 1 candidates.

Differences between candidates'  $G_4$  scores for the same grade in different subjects are notable. For instance, while a candidate with a  $G_4$  score of 77.0 would likely be awarded a Grade 1 in Graphical Communication, a candidate with the same  $G_4$  score would more likely be awarded a Grade 2 in most of the other subjects. However, small differences are expected, especially given the non-standard measure of candidates' general ability. Moreover, although a general placement of subjects from more severely graded to less severely graded can be observed in Figure 3 above, the order of subjects is not maintained for each grade. For instance, while Chemistry, Biology and Environmental Studies have the highest mean candidates'  $G_4$  score for Grade 1, the order changes to Environmental Studies, Chemistry and Biology for Grade 2. Additionally, the relative difficulty of subjects such as Italian, Religion, Computer Studies, Accounts, and German varies in a rather erratic manner per grade as can be confirmed from a close examination of Figure 3.

### *One Way ANOVA*

A finer analysis can be obtained by considering the  $G_4$  scores of candidates obtaining Grades 1 to 3 and those obtaining Grades 5 to 7 separately in the various subjects. This is achieved by comparing the mean scores through a one-way ANOVA test. Fisher's least significant difference (LSD) post-hoc tests are run for the one-way ANOVA.

Table IV presents the number of candidates who achieved Grades 1-3 in the various subjects (N), their mean  $G_4$  score, the standard error, and the range of marks. The box plot that follows illustrates this information (Figure 4).

Table IV: Statistical Information about the General Score obtained by Candidates who achieved Grades 1-3 in the various subjects

<b>Subject</b>	<b>N</b>	<b>Mean</b>	<b>Standard Error</b>	<b>Range</b>
<b>Chemistry</b>	353	75.2	0.33	34.8
<b>Biology</b>	390	74.1	0.32	32.8
<b>Accounts</b>	185	73.5	0.47	44.3
<b>Environmental Studies</b>	392	72.8	0.35	44.8
<b>English Literature</b>	721	72.6	0.28	46.6
<b>Social Studies</b>	149	72.3	0.79	44
<b>French</b>	493	72.2	0.35	54.9
<b>Business Studies</b>	77	70.7	0.81	37.4
<b>German</b>	173	70.6	0.74	51.3
<b>Religion</b>	1045	69.9	0.27	51.1
<b>Computer Studies</b>	329	69.7	0.45	49.4
<b>Graphical Communication</b>	166	68.9	0.65	37.6
<b>Spanish</b>	72	68	1.23	55.3
<b>Italian</b>	518	67.1	0.55	71.5
<b>Physical Education</b>	103	64.2	0.96	44.6
<b>Design and Technology</b>	64	63.1	1.07	39.8
<b>Art</b>	144	62.9	1.06	63.7
<b>Home Economics</b>	252	60.6	0.59	58.2

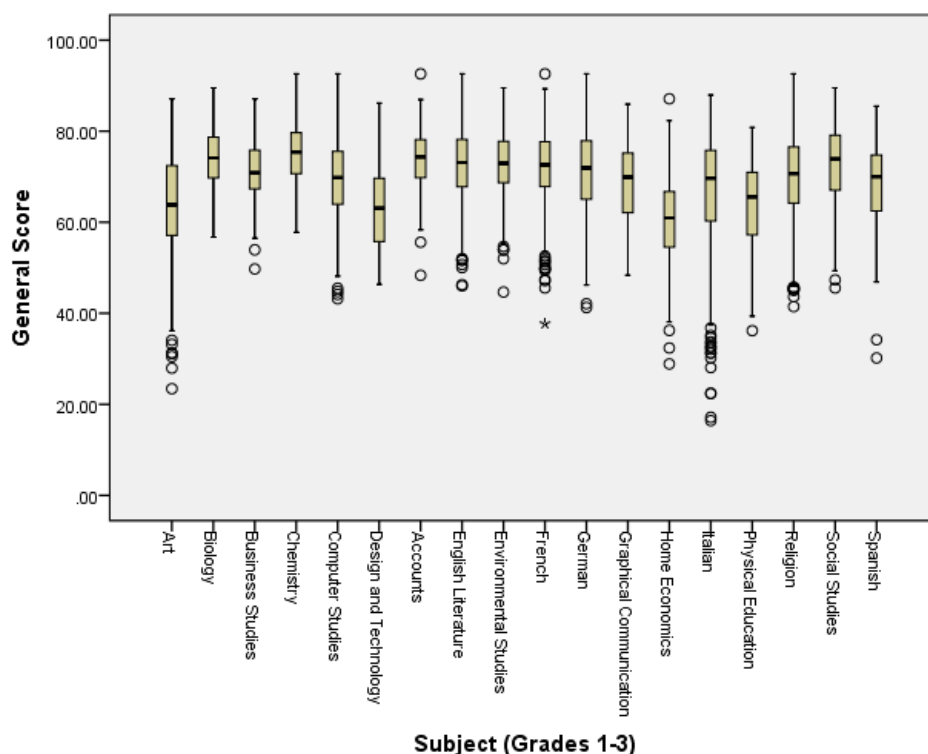


Figure 4: Box plot for the general score of candidates obtaining Grades 1-3 in the various subjects

The difference between the mean score of Home Economics Grade 1-3 candidates is statistically significantly lower than that for candidates in all other subjects. The mean differences for Art, Design and Technology, and Physical Education Grade 1-3 candidates are not statistically significant. However, for all these three subjects, the mean general score of Grades 1-3 candidates is statistically significantly lower than that for candidates from all other subjects (except Home Economics).

Besides the subjects discussed in the former paragraph, the mean score for Italian Grades 1-3 candidates is lower than that for all other subjects except Spanish, where this difference is not statistically significant. However, the difference between the mean score for Spanish Grades 1-3 candidates is not statistically significantly different than that for Business Studies, Computer Studies, Graphical Communication, and Religious Knowledge. The mean  $G_4$  score of German Grades 1-3 candidates is not statistically significantly different than that for Business Studies, Computer Studies, Graphical Communication, Religious Knowledge, and Social Studies. This inability to find a clear divide

between subjects persisted and the mean  $G_4$  score of candidates obtaining Grades 1-3 in Chemistry, the subject with the highest mean  $G_4$  score, is not statistically significantly different from that for Biology. This suggests that, for Grades 1-3, it is difficult to clearly divide these subjects into groups of candidates with different aptitudes.

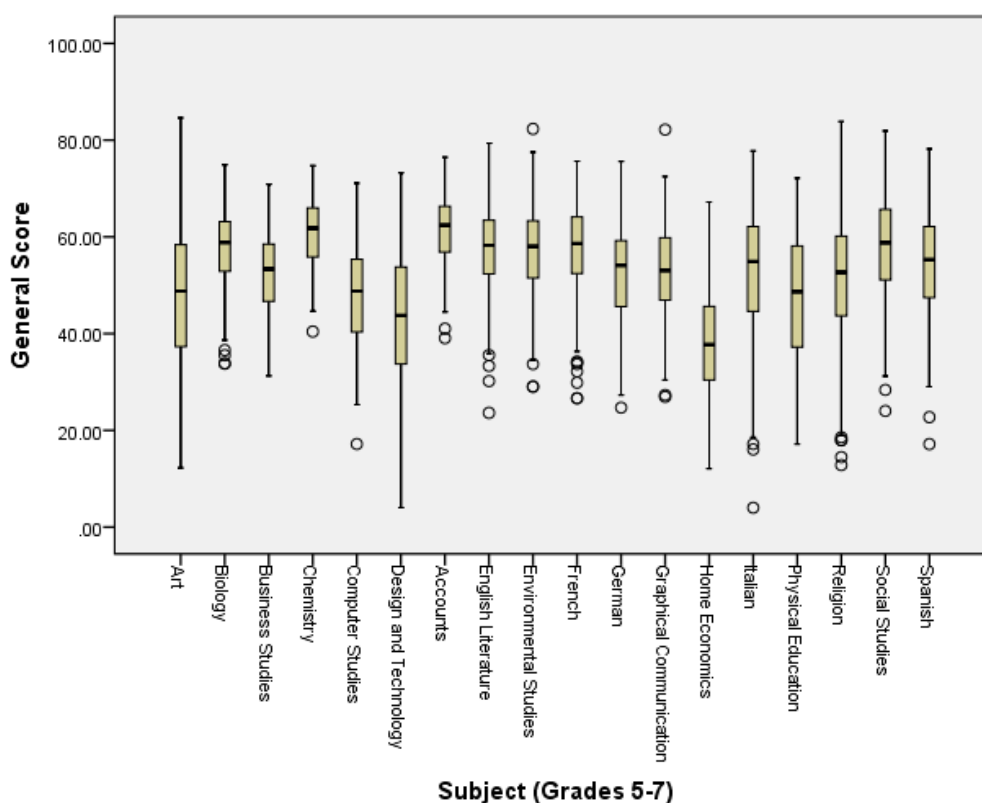


Figure 5: Box plot for the  $G_4$  scores of candidates obtaining Grades 5-7 in the various subjects

For the mean general score of candidates obtaining Grades 1-3 in the selected subjects, there is a clear difference between the mean general score of candidates obtaining Grades 1-3 in Home Economics and all the other subjects. The differences between the general scores of Art, Design and Technology and Physical Education are not statistically significant however their scores are different from the mean general score of other subjects. Thus, these three subjects can be grouped together in terms of candidates'  $G_4$  scores. For the remainder of the subjects there is no clear point where a clear break in terms of candidates' mean  $G_4$  scores can be made. These divisions have been marked on Table IV using dotted lines.

A similar analysis is presented in Table V and Figure 5 for the mean  $G_4$  scores of candidates obtaining Grades 5-7 in the various subjects.

Table V: Statistical Information about the General Score ( $G_4$ ) obtained by Candidates obtaining Grades 5-7 in the various subjects

<b>Subject</b>	<b>N</b>	<b>Mean</b>	<b>Standard Error</b>	<b>Range</b>
<b>Chemistry</b>	131	61	0.63	34.3
<b>Accounts</b>	82	61	0.9	37.4
<b>Biology</b>	193	57.6	0.59	41.1
<b>English Literature</b>	463	57.6	0.41	55.8
<b>Social Studies</b>	200	57.5	0.77	57.9
<b>French</b>	221	57.4	0.67	49
<b>Environmental Studies</b>	339	56	0.49	53.3
<b>Spanish</b>	71	54.4	1.33	61.1
<b>Graphical Communication</b>	132	52.8	0.9	55.3
<b>German</b>	80	52.6	1.27	50.9
<b>Italian</b>	418	52.6	0.63	73.8
<b>Business Studies</b>	112	52.4	0.83	39.6
<b>Religion</b>	718	51.7	0.44	71
<b>Art</b>	135	48.5	1.24	72.4
<b>Computer Studies</b>	165	47.8	0.81	53.9
<b>Physical Education</b>	65	46.1	1.77	55
<b>Design &amp; Technology</b>	101	43.5	1.31	69.3
<b>Home Economics</b>	118	38.5	1.06	55.1



As in the case of candidates obtaining Grades 1-3, the average  $G_4$  score of candidates obtaining Grades 5-7 in Home Economics is statistically significantly lower than the mean  $G_4$  score in any other subject. The mean  $G_4$  score for candidates obtaining Grade 5-7 in Design & Technology is lower than that for all subjects besides Home Economics, where the score is higher, and Physical Education, where the difference of 2.6% is not statistically significant. The differences in mean  $G_4$  scores between candidates obtaining Grades 5-7 in Art, Physical Education, and Computer Studies are not statistically significant, although these are statistically significantly lower than the candidates'  $G_4$  score for all other subjects. As in the case of the analysis for Grades 1-3, these subjects could not be placed into separate groups.

Therefore, the identical analyses carried out for the general score of candidates obtaining Grades 1-3 and that conducted for the candidates obtaining Grades 5-7 suggest the same conclusion: there is no clear point where a distinction between the mean general score of the various subjects can be clearly made with the exception of Home Economics and another separate group of subjects whose composition varies slightly for the different grade ranges.

## **Conclusion**

The belief that certain subjects are more challenging than others is widespread. This brief research study aims to shed light on these perceptions in order to assess whether these are factual and, if they are, their extent. The research is based on a number of assumptions which need to be clearly spelt out in order to consider results within their context. The most obvious of these will be discussed below before summarising the interpretation of results.

### *A General Ability*

It is assumed that there is a candidate general ability score that can be obtained by calculating a candidate's mean mark in four core subjects: English Language, Maltese, Mathematics and Physics. The research excludes candidates who did not sit for exams in these four subjects in the Main 2016 session. This assumption may be criticised as it may exclude candidates who excel in other forms of intelligence identified by Gardner (2004). Nevertheless, research by Rindermann (2007) suggests that cognitive tests essentially measure the same construct as candidates who do well in one academic area are likely to do well in another, with correlation being rather strong. Ochanji (2000) argues that while teachers adopt Gardner's theory of multiple intelligences, controlled, norm-referenced assessment gives value to particular general traits.

The values for correlation between Maltese, English Language, Mathematics and Physics presented in Table 2 are medium to high in strength. Moreover, data presented in Figure 3 suggest that a candidate's mean  $G_4$  score as measured for the purpose of this study follows the order of Grades as obtained by candidates in the various subjects: Grade 1 candidates in a subject have a higher mean  $G_4$  score than those obtaining Grade 2, who have a higher  $G_4$  score than those obtaining Grade 3 and so on. This was the case in all instances with one minor exception in Art, where Grade 2 candidates have a higher mean general score than those obtaining Grade 1. Exceptions were noted in the case of Grade U candidates in a number of instances, however this can be explained by the fact that Paper IIA candidates will get an Unclassified Grade if they do not obtain at least Grade 5 but they may have a much higher  $G_4$  score than Paper IIB candidates who do not manage to obtain a Grade 7.

The case of Art, however, suggests that there are at least a few instances where the assumption of a general ability score which predicts candidates' achievement in all subjects does not hold. This could be due to a number of reasons, including the theory of multiple intelligences. Another factor is student interest in the subject which, though affected by their socialisation, does affect their test results (Connor & Vargyas, 1992). Such differences were also observed by Coe (2008), as previously indicated.

This discrepancy exposes another assumption adopted when using candidates' test results: It is assumed that these results are a true indication of a candidates' ability. It is assumed that candidates sitting for a national examination are doing their best to perform well in that assessment. Although this sounds obvious, it excludes candidates sitting for examinations because of family pressure, candidates who opt not to study certain subjects well to focus on others, and those candidates sitting for an assessment in a subject which they did not study at enough depth in school. Although there might be other conditions which affect performance, like candidates' stress levels during examinations, some of these only do so in the assessment of specific subjects thus decreasing the link between measured general ability and performance in such subjects.

## Interpretation of Results

Through separate analyses of Grades 1-3 and Grades 5-7, subjects were divided into three groups. In both cases, Home Economics was classified alone since the candidates' mean  $G_4$  score in the subject was lower than that for all other subjects. Art, Physical Education and Design and Technology were also grouped together in both analyses, although in the case of candidates obtaining Grades 5-7, Computer Studies was added to this group of subjects. The difference between the candidates' mean general score was not statistically significant between these subjects. However, it was higher than that for Home Economics, but lower than that for all the other subjects. The rest of the subjects were grouped together as there was no clear point where a distinction between one subject and the next in terms of mean  $G_4$  score could be made.

While this study has a number of limitations, such as its focus being limited to the Main 2016 examination session, it seems that candidate general academic ability as measured in this study is a good predictor of the grade obtained in most subjects. It is interesting to note, however, that Home Economics and Design and Technology are subjects with a coursework component which contributes 30% and 50% respectively to the final mark. This component is marked by the candidates' teachers and it is only moderated by the MATSEC marking panels. On the other hand, P.E. and Art are subjects which have practical non-written assessments. These factors might contribute to candidates with a lower  $G_4$  score, which is measured on cognitive criteria, opting for and succeeding to obtain high grades in these subjects.

## References

- Barbara, A., Muscat, D., & Zahra, G. J. (2010). *The SEC Chemistry Syllabus: A Study and Proposed Amendments*. Unpublished B.Ed. (Hons.) dissertation, Faculty of Education, University of Malta.
- Childs, P. E. & Sheehan, M. (2009). What is difficult about Chemistry?: An Irish perspective. *Chemistry Education Research and Practice*, 10(3), 204-218.
- Coe, R. (2008). Comparability of GCSE examinations in different subjects: an application of the Rasch model. *Oxford Review of Education*, 34(5), 609-636. doi: 10.1080/03054980801970312
- Connor, K. & Vargyas, E. J. (1992). The Legal Implications of Gender Bias in Standardized Testing. *Berkeley Journal of Gender, Law & Justice*, 7(1), 13-89. Retrieved from <http://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=1063&context=bglj>

- Crippen, K. J. & Brookes, D. W. (2008). Applying cognitive theory to chemistry instruction: the case for worked examples. *Chemistry Education Research and Practice*, 10(1), 35-41.
- Cutajar, J. A. (1999). Gender, Ethnicity and Education in Malta, *Convergence*, 32(1-4), 70-82.
- Fitz-Gibbon, C. T. & Vincent, L. (1997). Difficulties regarding subject difficulties: developing reasonable explanations for observable data. *Oxford Review of Education*, 23(3), 291-298. doi: 10.1080/0305498970230302
- Gardner, H. (2004). How Education Changes: Considerations of History, Science and Values. In M. M. Suárez-Orozco & D. B. Qin-Hilliard (Eds.), *Globalization: Culture and Education in the new Millennium* (pp. 235-258). University of California Press and Ross Institute.
- Gardner, J. (2013). The public understanding of error in educational assessment. *Oxford Review of Education*, 39(1), 72-92. doi: 10.1080/03054985.2012.760290
- Good, F. J. & Cresswell, M. J. (1988). Grade awarding judgements in differentiated examinations. *British Educational Research Journal*, 14(3), 263-281. Retrieved from [http://www.jstor.org/stable/pdf/1500982.pdf?\\_=1469689091052](http://www.jstor.org/stable/pdf/1500982.pdf?_=1469689091052)
- MATSEC Support Unit, University of Malta (2015). *Grade Awarding: Guidelines*. Retrieved from <http://www.um.edu.mt/matsec/examiners>
- Newton, P. (2007, August). Techniques for monitoring the comparability of examination standards. Paper presented at the International Association for Educational Assessment 33rd Annual Conference, 16-21 September 2007, Baku, Azerbaijan.
- Newton, P. E. (2012). Making sense of decades of debate on inter-subject comparability in England. *Assessment in Education: Principles, Policy & Practice*, 19(2), 251-273. doi: 10.1080/0969594X.2011.563357
- Ochanji, M. (2000). Rethinking the role of the science teacher. *The Science Teacher*, 67(5), 24-27.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-Tests across nations. *European Journal of Personality*, 21(5), 667-706. doi: 10.1002/per.634
- Wilson, J. (1996). Revising the comprehensive ideal. *British Journal of Educational Studies*, 44(4), 426-437. Retrieved from <http://www.jstor.org/discover/10.2307/3121913?uid=3738632&uid=2&uid=4&sid=21103879493233>

