

**IMFM**

INSTITUTE OF MATHEMATICS, PHYSICS AND MECHANICS  
JADRANSKA 19, 1000 LJUBLJANA, SLOVENIA

**Preprint series**

**Vol. 50 (2012), 1177**

ISSN 2232-2094

**REMARKS ON EQUATING  
OF GRADES AT BASIC AND  
HIGHER LEVEL OF  
MATHEMATICS  
ACHIEVEMENT**

Janez Žerovnik

Ljubljana, June 14, 2012

# Remarks on Equating of Grades at Basic and Higher Level of Mathematics Achievement

Janez Žerovnik<sup>1</sup>

## Abstract

It is shown that application of the Rasch model for equating grades at Basic and Higher Level of Mathematics at General Matura exam in [Hauptman, Metodološki zvezki 7 (2010) 167-181.] is not justified.

## 1 Introduction

In Slovene General Matura, Mathematics is one of the obligatory subjects taken by all candidates. The candidates can choose between Basic Level and Higher Level exam. Besides being the required school-leaving exam, General Matura is a general admission requirement for University studies. Furthermore, the score at Matura is the main criteria for admission to some popular faculty courses with limited number of admissions. It is not surprising that the discussion of various aspects of Matura is extensive, from political in popular media to educational and theoretical in specialized journals. Although exam in Mathematics is one of the obligatory exams, only a few research papers are focusing on its theoretical analysis [4, 1]. Even worse, the present author is seriously sceptic about the methods and even more about the interpretation and conclusions in [1]. This will be clarified in more detail in this short note.

## 2 Structure of the exams

Both Basic and Higher Level exam in Mathematics usually include the same Part 1 of the exam, with only a minor difference that the time allowed to solve the problems of Part 1 is 90 minutes for Higher level candidates and 120 minutes for Basic level candidates. The second part of the exam (Part 2) is applied only to the Higher Level's candidates. Part 1 amounts to 80% of the total points at Basic Level, and 53.3% of total points at Higher Level. Higher Levels candidates get other 26.7% of points in Part 2. Oral part of the exam represents 20% of the grades at both levels, but of course some of the questions at the Higher level exam are more difficult and are taken from larger list than questions at the Basic level. The achievement in the General Matura is expressed by the classic five-grade scale including grades Insufficient (1), Sufficient (2), Good (3), Very Good (4) and

<sup>1</sup> FME, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia; janez.zerovnik@imfm.uni-lj.si

Excellent (5). The achievement in subjects taken at Higher Level is expressed on the scale from 1 to 8, which can be translated into traditional as Insufficient (1), Sufficient (2), Good (3,4), Very Good (5,6) and Excellent (7,8). Candidates pass the exam in each subject (and at each Level) if they at least get the grade 2. The conversion of points into grades (i.e. using limits of intervals of points (which are called boundaries) leading to grades) at each subject and on each Level is set independently. A proposal for the boundaries (i.e. how many percentage points are required in any individual subject for each grade) is prepared by the Subject testing Committees for the Matura, following the combination of relative and absolute criterion of assessment.

Part 1 consists of 12 problems (solutions of them bring up to 5,6,7 or 8 points), and Part 2 has 3 problems each having 3 or 4 subproblems, that may or may not be independent. Maximal total score for Part 1 is 80 points, and for Part 2 it is 40 points. The exams until the year 2004 had the same structure (i.e. Part 1, Part 2, and oral part) but the conversion of points into grades was slightly different. In short, the weight of Part 1 and Part 2 was 40% and the conversion of points into grades was more complicated and at the same time did not allow to assign grade 7. The structure of Part 1 and Part 2 of the exam however did not change which means that the observations of [4] may be relevant for the present exam as well. It is important to know that the majority of Part 1 problems are designed to test the basic knowledge, i.e. applying basic methods and understanding basic notions, while in Part 2 the solutions involve somewhat higher level of understanding and mastering the subject. Hence the intention is to measure different type of knowledge in Part 2 than in Part 1, Higher level naturally meant to be an upgrade of the Basic level.

Hence all parts of the exam are measuring knowledge of mathematics, but on different levels. The results on Part 1 can be used to compare the two populations (taking the Basic and Higher Level exam), the oral part of the exam is different for the two populations, and Part 2 is taken only by the candidates at higher level.

### 3 Questions of interest

There are many interesting general and specific questions related to the General Matura exams in Mathematics. For example, it is believed that it is important for the future progress of national economy that a larger proportion of students enters natural science and technical faculties, which implies that it is desirable to have a larger proportion of population entering the General Matura exam at Higher level. In particular, because this assumes that these students are better prepared for the studies as higher level exam covers more material implying more effort for good preparation is needed. However, although very unlikely, it is possible that, with the same knowledge, a candidate would receive lower grade at Higher level than he/she would get at the Basic level exam. How likely it is? In view of above, can we say that it is profitable to prepare for the Higher level exam without too much risk to receive even lower grade than at the simple Basic level exam? On the other hand, a very general question of interest is simply to compare the grades a hypothetical candidate is expected to achieve at the two levels of exam.

An attempt to equate the grades was explained in [1] concluding that there are certain discrepancies observed on the data from General Matura 2008. For example, the author of [1] claims that the candidates receiving grade 5 at the Basic Level exam should receive

grades 6,7, or 8, and on the other hand, that the candidates who received grade 5 at Higher Level exam should receive lower grade. If correct, this conclusions would impose serious doubts on the validity of grades at the General Matura exams in Mathematics, for the year 2008 for which the data were analysed and most probably for all other years because the stucture of the exam and the design of exams' problems was not essentially different. This claims (as far as the author of this note is aware) did not trigger any reaction. A reason may be that the conclusions are simply unreasonable. However, we believe that it is necessary to answer to such claims by questioning the methods which lead to this problematic conclusions which is done in the next section.

#### 4 Methods of analysis used in [1]

There are many statistical methods that may be applied to approach the question(s) mentioned above. As the problems in Part 1 of the exam on Higher level are the same as problems on Basic level, it may be natural to think of comparing the two exams using the item response theory (for introduction see, for example [2]). Item response theory overcomes the limitation of the Clasical test theory where the examinee characteristics cannot be separated from test characteristics. It is an important tool for design of psychological tests which can in principle be free both from test item and examinee bias. The method involves dynamic generation of later questions (items) based on scores on previous ones. As the General Matura exam is written only once and at the same time for the whole population, there is no need for such complicated test design. On the other hand, at the first sight it seems possible to use the item response theory for comparing the measured scores on the two levels of exam.

There are two basic postulates in the item response theory: (1) the performance of an examinee on a test item can be explained or predicted by a set of factors called "abilities". It is assumed that all items of the test measure only one ability and that various items of the test are statistically independent. (2) the relationship between examinees' item performance and the abilities underlying item performance can be described by a monotonically increasing function called item characteristic curve. This curve is interpreted as probability of correct answer to a YES/NO query as a function of person's ability expressed in logits. Persons logit is  $\ln(P_i/(1 - P_i))$  where  $P_i$  is the proportion of items that the examinee correctly answered. Here we meet the first serious problem of the analysis made in [1], where it is said that the Rasch's model is used for analysis. The Rasch model is often considered to be an item response theory model, namely the one-parameter model, although it includes a completely different approach to conceptualizing the relationship between data and the theory. While a generalization of the Rasch's model from dichotome items to polytome items (allowing for example answers strong disagree (SD), diasgree (D), agree (A), and strong agree (SA)) is well known, it unclear how to apply the model to tests with items for which partial solutions are possible. For example, in the matura exam Part 1 the problems' solutions are evaluated by 0 to maximum of 5,6,7, or 8 points, the proportions and the probability of correct answer might be replaced by some quantity related to expected score. Unfortunately, in [1] there is no definition of the model used and no explanation how the scores on items are treated, even worse, there is no reference to the model used besides naming it Rasch model. Furthermore, in Part

2, there are three problems with 3 to 4 subproblems each. It looks like in [1] the subproblems are taken as the items, but it is not explained. However, the subproblems are very likely not statistically independent contradicting one of the basic postulates of the item response theory. Even if one believes that extensions of basic theory that are not explained were sound and correctly applied in [1], there is another serious concern on the basic assumptions.

The psychological quantity that is measured in the General Matura exams in Mathematics can conditionally be called "mathematical knowledge" or "mathematical ability". Although the performance is finally evaluated as a grade (i.e. a one dimensional quantity), it is clear that there are many different sometimes loosely correlated abilities that constitute the "mathematical knowledge" or "mathematical ability". The exams are therefore designed in a way to measure various type of knowledge or abilities of candidates. Besides differences among the topics (for example skills in algebra, geometry, or calculus) there is a variety of knowledge levels tested (points may be received for knowing a formula or a theorem, for applying them, for expression manipulation, for numerical calculations, etc.). This differences are in particular emphasized between majority of items in Part 1 comparing to Part 2 of the exam. The assumption that both parts of the exam measure the same ability is therefore very questionable, and should have been argued more carefully and/or investigated more closely. For example already [4] observes that the two parts of the exam (in year 1998) measure somewhat different abilities: "Part 2 is not just an extension of Part 1, but rather measures other type or quality of knowledge than Part 1" (in Slovene: "Pola 2 ni bila le podaljšek pole 1, ampak je merila drugo vrsto ali kakovost znanja kot pola 1"). Finally, it is not clear how the oral part giving 20% of the grade is taken into account when equating the grades. The results of this part may be of poor quality. It has been observed that the oral part of the exam, which is the internal part of the exam, is unreliable in the sense that the comparison of scores of internal and external parts shows considerable differences (i.e. low correlation) in general and even great inconsistencies at some schools [5]. In conclusion, most probably, the very basic assumption of the Rasch's model [3], unidimensionality, is not valid for the data studied and hence the use of Rasch's model is not justified.

## 5 Conclusions

As indicated above, the method and in particular the assumptions used in the analysis [1] are highly questionable. Therefore it is important that the conclusions drawn in [1] are disregarded. On the other hand, the questions raised are interesting both from theoretical and practical viewpoint, in particular because of the importance of the General Matura scores for whole generations of candidates. Further research of here mentioned and related questions is needed.

## References

- [1] Hauptman, A. (2010): Equating of Grades at Basic and Higher Level of Mathematics Achievement. *Metodološki zvezki*, **7**,167-181.

- [2] Henard, D. (2000): Item Response Theory. in: Grimm, L.G. and Yarnold, P.R., eds.: *Reading and Understanding more Multivariate Statistics*. American Psychological Association, Washington, 67-97.
- [3] Linacre, j. M. (2011): Rasch Measures and Unidimensionality, *Transactions of the Rasch Measurement*, **24**, 1310.
- [4] Poljanšek, A. (2000): Ustreznost vrednotenja znanja pri maturitetnem izpitu iz matematike. *Psihološka obzorja*, **9**, 69-78.
- [5] Zupanc, D., Bren M. (2010): Inflacija pri internem ocenjevanju v Sloveniji. *Sodobna pedagogika*, **3**, 208-228.