

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

Alenka Hauptman<sup>1</sup>

## Abstract

In Slovene General Matura, Mathematics is one of the compulsory subjects and it can be taken either at Basic or Higher Level of Achievement. Basic Level of Achievement is expressed by the classic five-grade scale from 1 to 5. Candidates at Higher Level of Achievement can get grades on scale from 1 to 8. Conversion of points into grades (i.e. getting points on tests and points at internal examination and then calculating those grades from the sum of points) on each Level is set independently, and we tried to find out if the same grade on each Level of Achievement corresponds to the same knowledge. Once grades are assigned they are used comparatively in selection procedures for admission to University.

Both Basic and Higher Level in Mathematics include the same Part 1 of the exam. 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 Level's candidates get other 26.7 % of points in Part 2. Oral part of the exam represents 20 % of the grades at both Levels.

In this paper we show discrepancy between knowledge within the same grades for candidates at Basic and Higher Level of Achievement on an example of a Mathematics exam from General Matura 2008<sup>2</sup>. Rasch model within Item Response Theory framework was used to place item difficulties on common scale and the comparability of grade conversion on both Basic and Higher Level of Achievement was explored. The results show interesting differences in knowledge of candidates with the same grade at Basic and Higher Level of Achievement.

## 1 Introduction

This paper is focused on the discrepancy of grades at two different Levels of achievement. Mathematics and some foreign languages in Slovene General Matura can be taken either at Basic or Higher Level of achievement (Bahovec, 2009). In this paper the focus is on Mathematics, where the differences between the two

---

<sup>1</sup> National Examinations Centre, Ob železnici 16, 1000 Ljubljana; alenka.hauptman@gmail.com

<sup>2</sup> Data from Slovenian General Matura were obtained through National Examinations Centre.

Levels are more explicit (Vogel, 2008). Since the conversion of points into grades (i.e. getting points on tests and points at internal examination and then calculating grades from the sum of points) on each Level is set independently, we tried to find out if the same grade on each Level of achievement corresponds to the same knowledge. We are interested in the problem of equating two levels of achievement practically, since Slovenian General Matura examinations are main high stakes exams in Slovenia and they represent the certificate of secondary education as well as enable the student to enter the University (Gabršček and Bethell, 1996), but also scientifically, even though the problem of equating two different levels is not completely new in the research areas (e.g. Poljanšek, 2000).

## 2 Slovene General Matura

Slovene General Matura is a school-leaving exam required for the completion of the secondary education and for university entrance (Bahovec, 2009). Thus the General Matura represents not only a final exam but also a continuation regulating the transition from secondary to tertiary education, as is the case of educational systems in many European countries (Gabršček, 1996). The General Matura is a national exam with equal conditions for all candidates: they take the exams simultaneously, following the same procedures and rules and in accordance with the same criteria of assessments.

A pass in the General Matura is a general admission requirement for any academic course and a minimal admission requirement for those academic courses having no limit as to the number of students (Bahovec, 2009).

Candidates sit the Matura in 5 subjects. Subjects of the compulsory part are Slovene (or in ethnically mixed areas, Hungarian or Italian), Mathematics and a modern foreign language (English, French, German, Italian, Russian or Spanish). Two subjects are optional, which candidates choose among a wide array of subjects (Gabršček and Bethell, 1996).

Mathematics and some foreign languages can be taken either at Basic or Higher Levels of achievement; candidates choose the Level some months before the exam is actually taken (Bahovec, 2009). Candidates choose the Level according to their interest and ability for each subject. Levels differ in their amount of required knowledge and the quality of knowledge (higher taxonomy goals) (Lorenčič, 1995). The Higher Level of achievement means a more demanding examination and not higher knowledge per se, because all candidates studied the same curriculum. Each candidate can take no more than two subjects at Higher Level. General Matura is held in two exam sessions, the Spring one and the Autumn one. The majority of the students take the Spring exam session, while mostly those who retake a failed exam or improve their result from the Spring session, take the Autumn exam session.

Each Matura exam typically includes the external part of the exam, prepared by subject experts on national level, and internal part of the exam, which includes the same national questions but asked and graded by teachers in the school of a student. The external part is in most cases written, whereas the internal part can be oral or represented by laboratory or course work. The final score in each Matura exam is the combined sum of points achieved by a candidate in individual parts of the exam (Vogel, 2008).

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 Excellent (5). The achievement in subjects taken at Higher Level is expressed on the scale from 1 to 8 (Bahovec, 2009). 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 procedure of defining boundaries is called setting the boundaries. 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.

### **3 Mathematics in Slovene General Matura**

As we already noted, Mathematics is one of the compulsory subjects in General Matura and it can be taken either at Basic or Higher Level of achievement.

The structure of both Levels is not equal. Basic Level's examination includes Part 1, which amounts to 80 % of the total points, and Oral part, which represents 20 % of the grade. Higher Level's examination includes Part 1, Part 2 and the Oral part. Part 1 of the Higher Level's exam is the same as Part 1 at Basic Level, but it amounts only 53.3 % of the total points. The second part of the exam (Part 2) is applied only to Higher Level's candidates and amounts to 26.7 % of total points. Oral part at Higher Level is partly the same as Oral part at Basic Level, but includes some additional questions. It also represents 20 % of the grade (Bahovec, 2009). Candidates choose the Level some months before the examination, so they can prepare for the Level they have chosen.

Reliability for Mathematics tests is through the years (from 1995 till 2008) relatively stable. Gutmann-Cronbach alpha is around .80 for both, Basic and Higher Level (Brešar, 1998; Vogel, 2008), which is an indication for relatively high reliability (Feldt and Brennan, 1993).

Even though some parts of the Mathematics exam on both Levels are completely the same, conversion of points into grades on each Level is set independently. The reason for that is that the grades are calculated from the raw

scores and transformed to 100 % scale. As a consequence each of the exams is treated like an independent exam.

Moreover, items on the Mathematics tests are not pretested, so we do not know what the difficulty of the test items is, even though that could help us with test equating. Based on these assumptions we can readily ask ourselves about grades' equating at both Levels.

Equating is a statistical process that is used to adjust scores on test versions so that scores on the test versions can be used interchangeably. Equating adjust for differences in difficulty among test versions that are built to be similar in difficulty and content (Kolen & Brennan, 2004). But in our case Basic and Higher Level in Mathematics cover the same content but differ in difficulty of items. The Higher Level includes some additional items that are more difficult than the items, included in Part 1 of the both exams, but cover the same content. Because each exam is treated like an independent exam, equating is not used in the Slovenian General Matura. In this paper we are trying to combine both tests on the same scale and compare the knowledge of students on both Levels.

Candidates who took exam at Basic Level did not take all the items, so the classical test theory was not the appropriate method to test the equivalency of Basic and Higher Level of achievement (McDonald, 1999). Instead of classical test theory we used the Rasch model, which can deal with missing values because the estimation methods do not require any imputations of missing data or case-wise or pair-wise omission (Bond and Fox, 2007). Rasch model is one of the more robustness models among the IRT models and is as such appropriate for this research.

Further in the paper we tried to find out if the same grade on each Level of achievement corresponds to the same knowledge.

## 4 The method

To explore what discrepancies of grades at Basic and Higher Level of Achievement are, we took the Mathematics exams from General Matura 2008. We have chosen data from Spring exam sessions, because most of candidates took the Matura exams in spring, and that the sample is most representative of the population of the candidates. Both Basic and Higher Level's candidates were chosen for the research (7.487 candidates at Basic Level and 1.583 candidates at Higher Level).

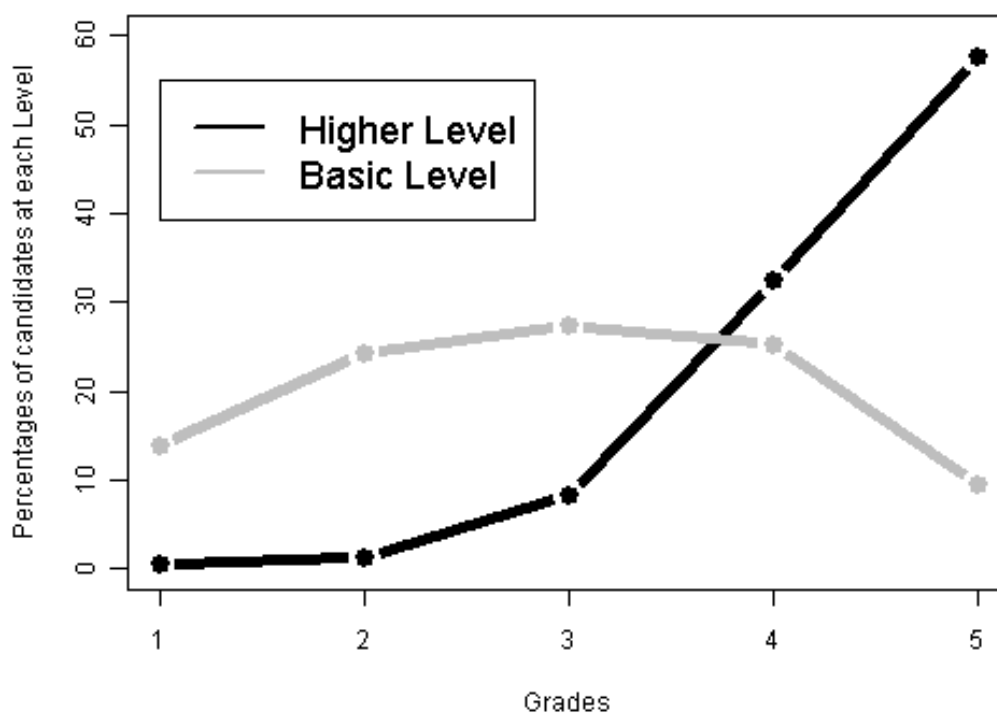
For the analysis we needed boundaries between grades, which were taken from the Annual Matura Report 2008 (Vogel, 2008), prepared by Mathematics testing Committee.

Two comparisons were employed: the classical test theory for comparison on Part 1 that is common to both Levels and the Rasch model to place item difficulties on common scale and the suitability of grade conversion on both Basic

and Higher Level of achievement was explored. Fit of the items to the Rasch model was also tested. For the Rasch's analysis specific software Rumm2020 was used using weighted maximum likelihood estimation method.

## 5 Results and discussion

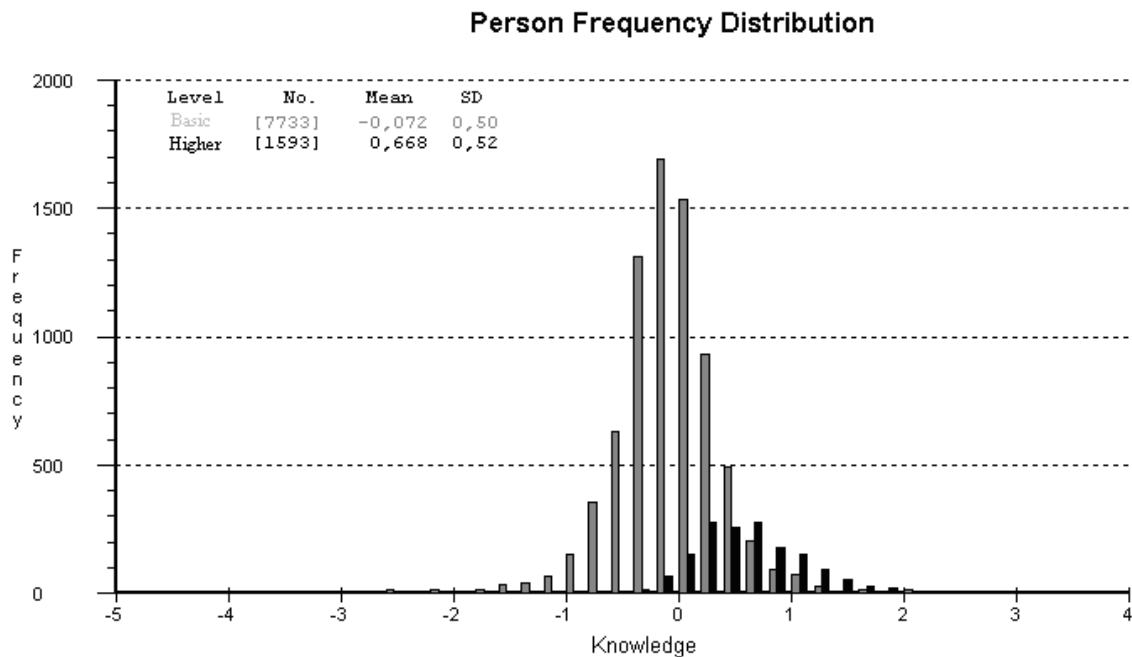
The comparison of candidates at Basic and Higher Level of achievement reveals that Higher Level's candidates are in average more competent in Mathematics. This is shown in Figure 1, which compares candidates at Basic and Higher Level, assuming they only took Part 1 of the Mathematics exam, which is the same for both Levels. Majority of Higher Level's candidates on General Matura 2008 would get grades 4 and 5, and the majority of Basic Level's candidates actually got grades 2, 3 and 4, even though some got a grade 5.



**Figure 1:** A comparison of the candidates at Basic and Higher Level in Mathematics (Part 1 of the exam).

Figure 2 shows a comparison of the candidates at Basic and Higher Level of achievement, assuming both groups took the Higher Level (Part 1 and Part 2 of the Mathematics exam). To do this, we used the Rasch model to place item difficulties

on a common scale (Basic and Higher Level together). Even though we didn't have any data of the Part 2 of the Basic Level's candidates, the Rasch model enables us to place all items on the same scale. Reliability of all items were calculated with the Person Separation Index (Andrich, 1988), which is based on average error variance. Person Separation Index is 0.90, which indicates high reliability (Andrich, Lyne, Sheridan and Luo, 2003).



**Figure 2:** Person frequency distribution.

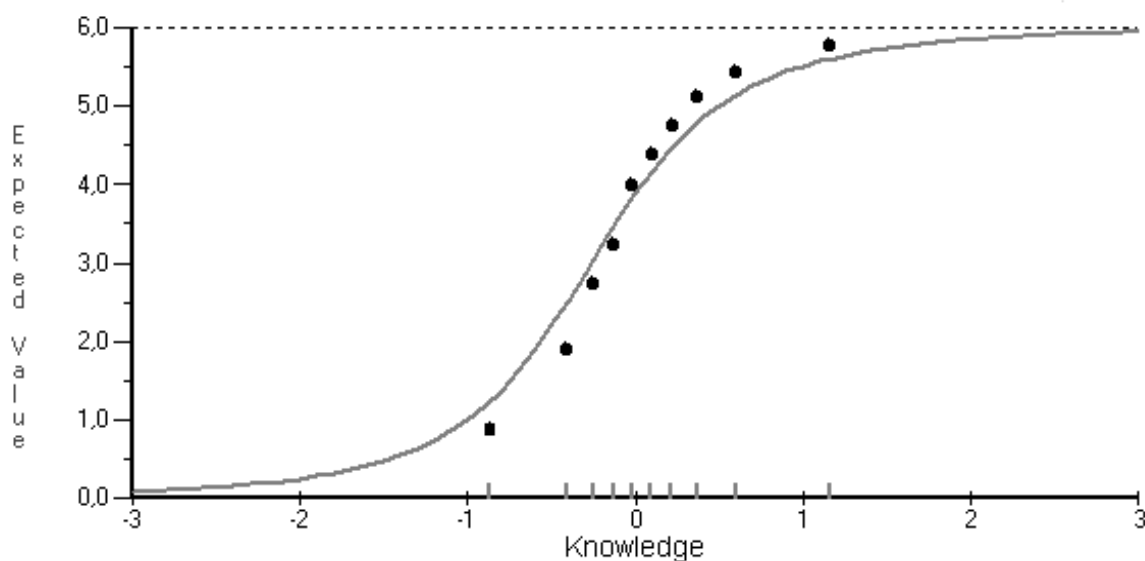
Figure 2 shows knowledge distributions of the candidates in Mathematics 2008, based on the actual data; the grey columns represent knowledge distribution of Basic Level's candidates and the black columns represent the knowledge distribution of the Higher Level's candidates.

The Basic Level's distribution ranges more widely than the Higher Level's distribution (from below -2 to above 2), which shows that the candidates vary a lot due to their knowledge. We can also see that some Basic Level's candidates have the knowledge of the best candidates at Higher Level, which reveals that these candidates could have better grades if they had taken a Higher Level in Mathematics.

Fit of the items to the Rasch model could be seen in the figures which show fitting of the actual data to the theoretical Rasch model, represented by item characteristic curves (ICC)<sup>3</sup>. Residuals of actual data (points on figures) from the

<sup>3</sup> Item characteristic curves (ICC) are charts of modelled response for participants on particular item.

theoretical ICC represent fit of the data to the Rasch model. Figure 3, where we can see an item which has the lowest fit to the model (Item 2), shows a statistically significant lack of fit of item 2 to the model ( $FitRes = -13.666$ ;  $\chi^2 = 0.000$ ). Although the deviation is statistically significant this is more due to the great sample size ( $N=9,070$ ) than actually important differences. Figure 3 shows that an Item 2 discriminates better than average item in a test but the differences are not great. All other items in a test show even better fit estimates; therefore, we conclude that all of the items show quite good fit to the Rasch model.



**Figure 3:** Item characteristic curve for Item 2. Expected value = the most likely score on the item.

In the next two figures (Figure 4 and 5) we can see the knowledge distributions for each grade at Basic and Higher Level respectively. Distributions were calculated from the actual data.

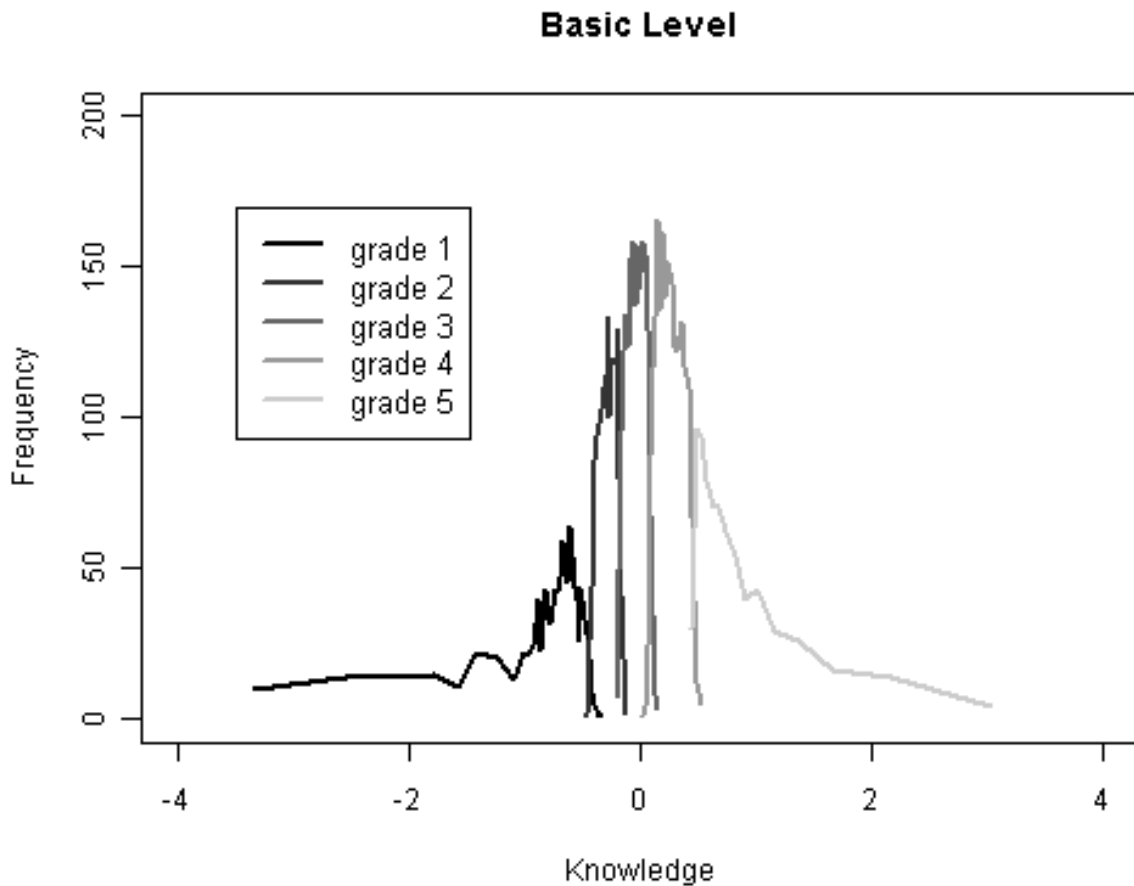
In both figures (Figure 4 and 5) knowledge distributions for each grade are marked with different shades of grey; for example: the lightest grey distribution represents the distribution of knowledge for the candidates with grade 5.

Grades boundaries could be defined between two neighbour knowledge distributions. For some knowledge distributions we can easily define the boundaries between two neighbouring grades, but some knowledge distributions overlap each other (more often at Higher Level), so it is harder to define boundaries between two neighbouring grades.

Based on these distributions (Figures 4 and 5), we tried to find out where on the dimension of mathematical knowledge are grades' boundaries. At this point we should mention that student's raw results are discrete, but the Rasch model transfers the scale of results into a continuous scale (called knowledge), so we had

to calculate the boundaries between grades with interpolation<sup>4</sup>. For that we compared knowledge distributions of each two neighbouring grades. Boundary between two neighbouring knowledge distributions was set between the last point scale of knowledge where frequency distribution of lower grade was higher than the frequency distribution of a higher grade and the first point on knowledge scale with a higher frequency of a higher grade than a lower grade. Boundaries are points in Figures 4 and 5 where each of two neighbouring knowledge distributions crosses each other.

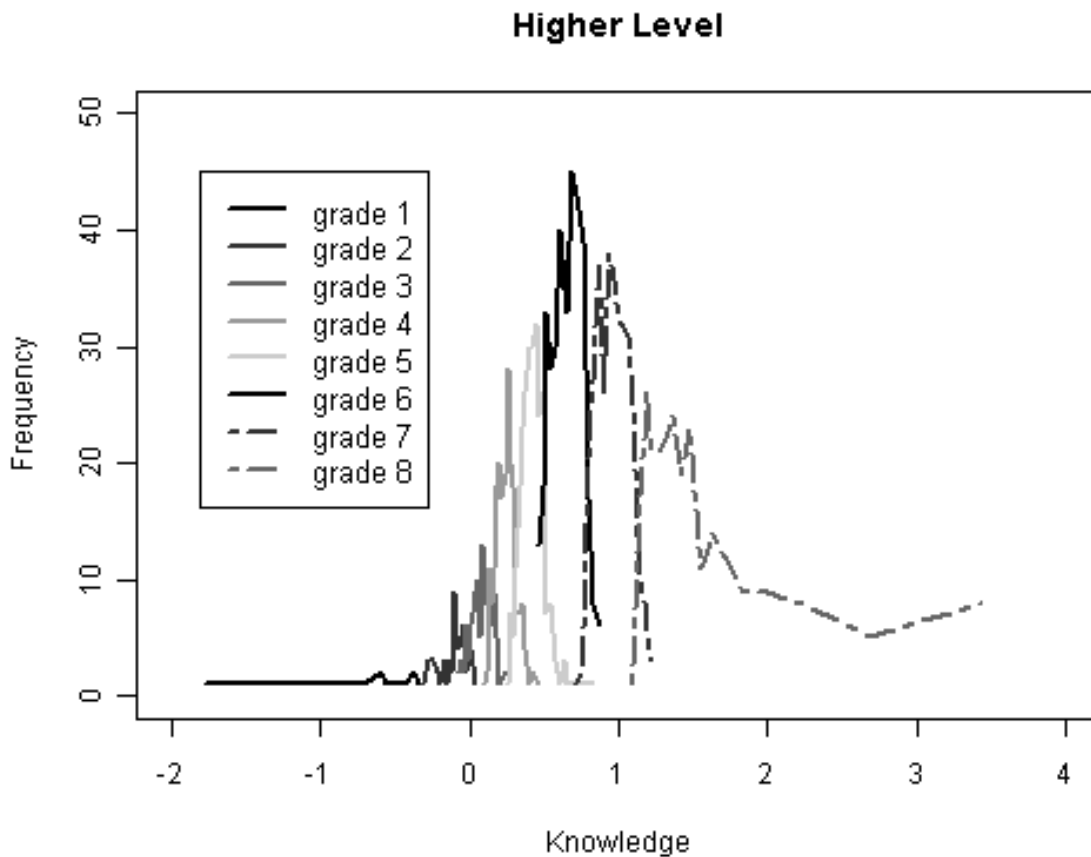
Boundaries between each of two neighbouring grades at Basic and Higher Levels are shown in Figure 6.



**Figure 4:** Knowledge distribution of Basic Level's candidates for every grade.

<sup>4</sup> A method of constructing new data points within the range of a discrete set of known data points.

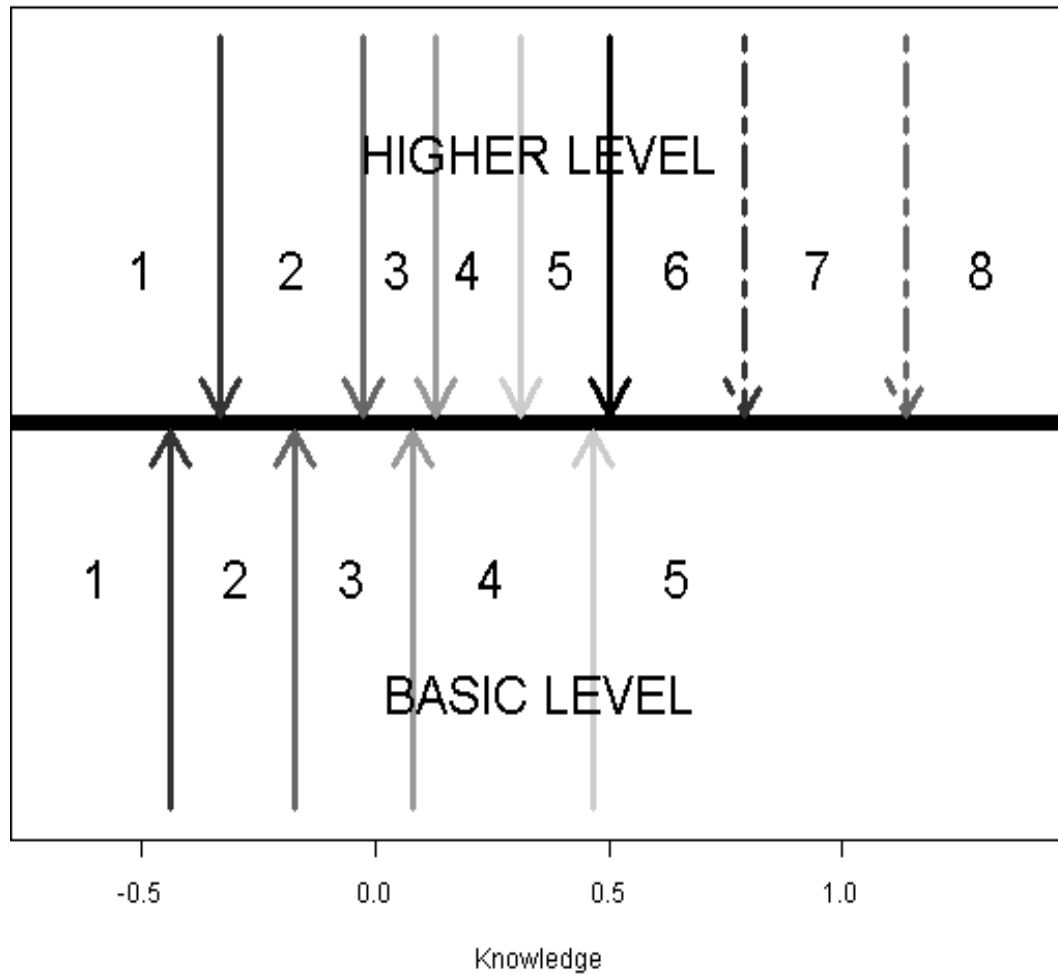




**Figure 5:** Knowledge distribution of Higher Level's candidates for every grade.

Figure 6 shows the grades' boundaries of Basic Level below the black row in the middle; and above the black row there are grades' boundaries of Higher Level. In this figure we can see that the boundaries at Higher Level are set at higher knowledge than boundaries at Basic Level. This indicates that the candidates at Higher Level must have a higher knowledge to get the same grade. But this is not the same for the boundary between grades 4 and 5, where candidates at Higher Level could have less knowledge than candidates at Basic Level to get grade 5.

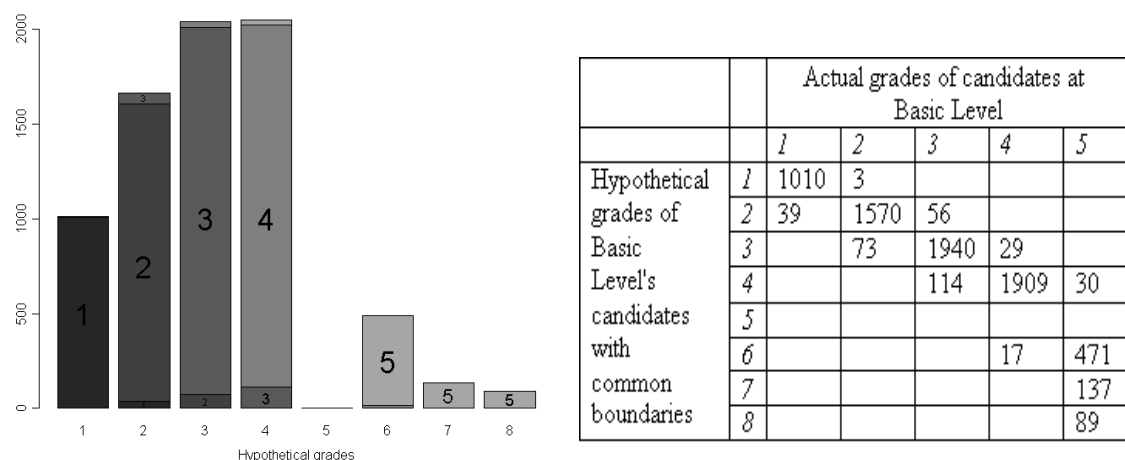
To evaluate the knowledge of the candidates regardless of the chosen Level, we proposed common boundaries based on those from both Levels; and to do this, we took the knowledge distributions for each grade of both Levels together and computed common boundaries. Because of the greater number of candidates at Basic Level the boundaries at lower grades would be more similar to boundaries at Basic Level, and at higher grades more similar to the boundaries at Higher Level. In the further analysis we were more interested in hypothetically equated grades, which could be one way to equate the grades of the candidates at Basic and Higher Level.



Note: Numbers from 1 to 8 represent grades.

**Figure 6:** Comparison of grade's boundaries at Basic and Higher Level.

In the next three figures (Figure 7, 8 and 9) we compare the actual and hypothetical grades. The actual grades (columns) are grades that candidates actually received and the hypothetical grades (x-axis) are grades calculated with the common boundaries.

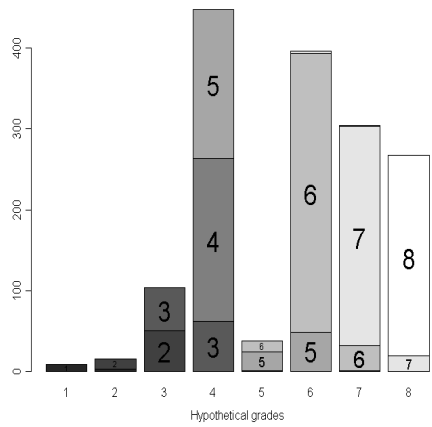


**Figure 7:** Comparison of the actual and hypothetical grades based on calculation with the common boundaries for the candidates at Basic Level.

Figure 7 represents the grades of candidates who had taken the Basic Level in the Mathematics tests. We can see that none of candidates would have a grade 5 with (hypothetically) common boundaries in the Mathematics tests. Instead of that, these candidates would have higher grades. That indicates that the candidates with knowledge for a grade 5 at Basic Level should take a Higher Level and probably would get a higher grade.

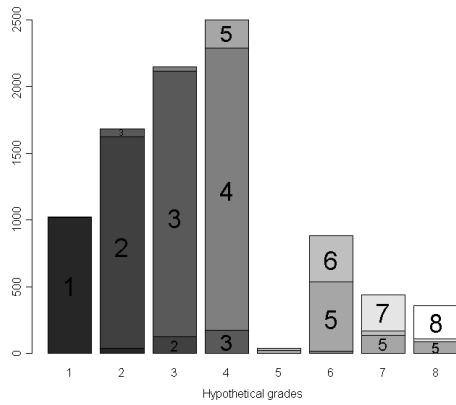
Figure 8 shows the grades of the candidates who took the Higher Level in the Mathematics tests. The Higher Level's candidates would have less low grades with common boundaries, but also less grades 5 on account to more frequent grade 4. That is because many candidates with the actual grade 5 at Higher Level have a lower knowledge than those at Basic Level.

Figure 9 represents grades of all candidates together (both, Basic and Higher Level's candidates). In this Figure we can see that hardly any one would have a grade 5 with common boundaries, and more would have higher grades (grades 6, 7 and 8). There is also more frequency in grade 4, which indicates that some lower grades would be higher with common boundaries and that some candidates with a grade 5 actually deserve this grade (as was already mentioned above).



		Actual grades of candidates at Higher Level							
		1	2	3	4	5	6	7	8
Hypothetical grades of Higher Level's candidates with common boundaries	1	9							
	2	3	13						
	3		51	53					
	4			62	202	184			
	5				1	24	13		
	6					49	344	3	
	7					1	31	271	1
	8							20	248

**Figure 8:** The comparison of actual and hypothetical grades based on calculation with the common boundaries for the candidates at Higher Level.



		Actual grades of candidates at both, Basic and Higher Level							
		1	2	3	4	5	6	7	8
Hypothetical grades of Basic and Higher Level's candidates with common boundaries	1	1019	3						
	2	42	1583	56					
	3		124	1993	29				
	4			176	2111	214			
	5				1	24	13		
	6				17	520	344	3	
	7					138	31	271	1
	8					89		20	248

**Figure 9:** Comparison of actual and hypothetical grades based on calculation with the common boundaries for candidates at both, Basic and Higher Level.

## 6 Summary and conclusions

In this paper we analysed the assumption that the same grade on each Level of achievement in Mathematics corresponds to the same knowledge and can be therefore used interchangeably.

The Higher Level's candidates in average have higher knowledge in Mathematics, but the knowledge of some candidates at Basic Level is also very high; some candidates at Basic Level even reach the knowledge of the best Higher Level's candidates, indicating that they would profit from choosing the Higher Level.

The analyses of grade's discrepancy at Basic and Higher Level of achievement show us that the same grade at each Level does not correspond to the same knowledge. Boundaries between two neighbouring grades are not the same at Basic and Higher Level; boundaries at Higher Level are higher for grades 2, 3 and 4 and lower for grade 5 than boundaries at Basic Level.

As we have seen in the results, the fact that we have two Levels of achievement in Mathematics presents some problems. First, the candidates with a grade 5 at Basic Level have to show a higher knowledge to get the grade 5 as Higher Level's candidates. Based on their knowledge they deserve higher grades. Second, it did happen that two candidates with the same knowledge but one at Basic and other at Higher Level got two markedly different grades – the one at Basic Level grade 5 and the other at Higher Level grade 8. Third, the candidates at Higher Level have to show a higher knowledge than the candidates at Basic Level to get grade 2, 3 and 4. This indicates that the candidates who took the Higher Level but perform badly, are somehow deprived.

For the unequal grades' boundaries at Basic and Higher Level it was hard to define common boundaries, which could be one way to equate grades of candidates at both Levels. The greatest problem was to define the boundary for grade 5. That is because the knowledge for this grade is not equal at Basic and Higher Level. Basic Level's candidates with grade 5 would at common boundaries mostly get higher grades, but Higher Level's candidates with grade 5 would, because of their lower knowledge, actually get a grade 4.

We can conclude that the same grade at different Levels of achievement does not represent the same knowledge. And the fact that the candidates choose the Level of examination in advance (and so could prepare for the chosen Level) does not justify grades dependent on the Level. The examinee's choice in Mathematics therefore introduces an error in measurement, which is congruent with the reports from other researchers (Bridgeman, Morgan and Wang, 1996).

There are more possibilities to equalize grades and knowledge. We can have one equal exam for all candidates. This is the case in Slovenian General Matura at Slovene language, where we have just one equal exam for all candidates with grades from 1 to 8. Another possibility is equating Basic and Higher Level exams,

having common boundaries. This is similar to our case, where we scaled all data on one scale and then looked where the boundaries would be. The third possibility is pre-testing the items, where we could have test characteristics and boundaries set in advance (Kolen and Brennan, 2004).

## Acknowledgement

The author would like to thank the National Examinations Centre for datasets provided. Opinions and interpretation, presented in this paper pertain to the author and are not affiliated with National Examinations Centre.

## References

- [1] Andrich, D. (1988): *Rasch Models for Measurement*. London: Sage Publications.
- [2] Andrich, D., Lyne, A., Sheridan, B., and Luo, G. (2003): RUMM2020. Perth: RUMM Laboratory.
- [3] Bahovec, I. (Ed.) (2009): *Maturitetni izpitni katalog za splošno maturo*. Ljubljana: Državni izpitni center.
- [4] Bond, T.G. and Fox, C.M. (2007): *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, (2<sup>nd</sup> ed.). Lawrence Erlbaum.
- [5] Brešar, F. (Ed.) (1998) : *Notranje vrednotenje mature: Poročilo*. Ljubljana: Državni izpitni center.
- [6] Bridgeman, B., Morgan, R., and Wang, M. (1996): *Choice Among Essay Tonics: Impact on Performance and Validity* (ETS Research Report 96-4). Princeton, New York: Educational Testing Service.
- [7] Feldt, L.S., and Brennan, R.L. (1993): Reliability. In Linn, R.L. (Ed.): *Educational Measurement* (3<sup>rd</sup> ed.) (pp.105-146). New York: American Council on Education, Oryx Press.
- [8] Gabršček, S. (1996): *Guide to Secondary School-Leaving Certificates in European Countries*. Cheltenham: UCAS.
- [9] Gabršček, S. and Bethell, G. (1996): *Matura examinations in Slovenia: case study of the introduction of an external examinations system for schools*. Ljubljana: National Examinations Centre
- [10] Kolen, M.J. and Brennan, R.L. (2004): *Test Equating, Scaling, and Linking: Methods and Practices*, (2<sup>nd</sup> ed.). New York: Springer.
- [11] Lorenčič, I. (1995): Ravni zahtevnosti. In Uršič, M. (Ed.): *Prvi strokovni posvet o maturi*. (33-35). Ljubljana: Republiški izpitni center.

- [12] McDonald, R.P. (1999): *Test Theory: A Unified Treatment*. London: Lawrence Erlbaum Associates.
- [13] Poljanšek, A. (2000): Ustreznost vrednotenja znanja pri maturitetnem izpitu iz matematike. *Psihološka obzorja*, **9**, 69-78.
- [14] R Development Core Team (2009): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- [15] Vogel, J. (Ed.) (2008): *Letno poročilo Splošna matura 2008*. Ljubljana: Državni izpitni center.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.