Jerneja Bone and Daniel Doz

Comparison of pupils' attainments between national assessments of knowledge and final school grades in mathematics

Abstract: National assessments of students' mathematical knowledge give educators, students, and policymakers some important feedback information about the quality of the educational system. However, this information is sometimes different from the one offered by mathematics teachers. Hence, it is important to consider both teachers' grades and students' attainments on standardised tests. This study attempted a longitudinal comparison of pupils' scores on the National Assessments of Knowledge in mathematics and their final mathematics grades at the end of their final year of primary school. It was found that the percentage of pupils with the final grade "excellent" is slightly on the rise, and that the percentage of pupils with the grade "sufficient" is in a slight decline. A high correlation was also found between the hypothetical National Assessment of Knowledge grade and the final mathematics grade received at the end of the pupils' final year of primary school.

Keywords: external assessment, grading, hypothetical grade, internal assessment, mathematics

Jerneja Bone, National Education Institute Slovenia, Organisational unit Nova Gorica, Erjavčeva 2, SI-5000 Nova Gorica, Slovenia; e-mail: jerneja.bone@zrss.si

Daniel Doz, assistant, University of Primorska, Faculty of Education, Cankarjeva 5, SI-5000 Koper, Slovenia; e-mail:daniel.doz@pef.upr.si

Introduction

Assessing students' knowledge is one of the most important factors in the learning process (Khadijeh and Amir 2015). The task of teachers, pupils, school policies and, last but not least, parents is to make meaningful use of the data obtained from internal and external assessments and interpret them accordingly (Kartianom and Mardapi 2017; Metsämuuronen and Ukkola 2022). While teachers and school policies make sense of the data or results obtained to improve or modify the teaching style and plan changes to the subject curricula (Felda 2018), pupils use the obtained data to plan their learning and advance their knowledge. According to Cankar et al. (2019), »[w]hen analysing and evaluating pupils' attainments, the school should ask itself two questions: how big the differences in pupils' attainments are at the school and between which groups, and where our attainments stand (e.g. at the level of Slovenia, compared with the previous school year, etc.)« (p. 9). Considering this statement, policymakers should ask themselves the following questions: how big are the differences in pupils' attainments between schools, between genders, and between the regions (municipalities) in the country? What do the trends of the longitudinal comparison of attainments indicate? Can the varying attainments be compared, and, if yes, how?

To manage data-based school policies and increase fairness, we must diagnose the target subgroups early on, collect the input and output data on these subgroups, conduct empirical evaluations, and select the appropriate measures. Such measures should be effective and (1) help teachers in the classroom, (2) help teachers run their schools, and (3) help in managing school policies at the state level (Wößmann 2008).

Significant information can be gained from external examinations or assessments (national or international). By analysing attainments, we can highlight different knowledge gaps, indicate the areas where lessons should be adapted, and contribute substantiated arguments to school policy reforms in individual subject areas. International assessments (e.g. PISA and TIMSS) have aided in the globalisation and comparability of mathematical knowledge. External and inter-

nal assessments also intertwine; they have many similarities and differences and rely on the principles of good assessment practice (validity, objectivity, sensitivity, reliability, and economy), which is why we expect a certain amount of harmony between them. Wyatt-Smith et al. (2014) suggested that we can offer young people better education when education policies and practices prioritise the improvement of learning and teaching.

At all stages of a mathematics lesson, teachers are oriented towards getting the pupils to advance their knowledge of mathematics and achieve the set objectives, which are laid down in the curriculum. Throughout the teaching process, pupils' knowledge is tested and graded, which is why their attainments are indicated by the grades obtained in internal knowledge assessments and their scores on external examinations or assessments. Their attainments are also demonstrated by their performance in mathematics competitions, though not all everyone attends them (Cankar et al. 2019).

Lindahl (2007), a Swedish study, compared the results of national tests with the teachers' assessment of the academic performance of final-year pupils (aged 16). The descriptive statistics revealed that the teachers usually gave more generous final grades at the end of schooling than justified by the results of the national tests. Especially for mathematics, a high percentage of pupils failed the national tests; however, the percentage of final fail grades was smaller. The author highlighted that teachers offer more generous final grades to pupils who are likely to fail the test.

In Slovenia, pupils' school grades and scores on the National Assessment of Knowledge are two sources of information on pupils' attainments that can be compared and studied from different angles. A norm-based comparison is sensible only if we convert the National Assessment of Knowledge scores into hypothetical grades awarded to the pupils (Felda 2018). The National Assessment of Knowledge enables us to monitor an individual or a group of pupils from a specific school, region, or the entire country over a longer period of time or in a given school year with regard to their knowledge of a select subject. To determine and assure the quality of classwork, it is essential that we monitor pupils' progress (OECD 2008).

A smaller survey conducted on a sample of pupils from one school aimed to determine how the final grade in mathematics correlates with the result from the external knowledge assessment and whether the grade given by the teacher differs considerably from the distribution of scores on the external knowledge assessment. The author established that, in the 1996/97 school year, the final grades were highly correlated ($r = 0.80$) with the external knowledge assessment (Smole 1999). In the 2005/06 school year, the correlation between the pupils' scores on the external knowledge assessment and the grades given by the teacher in mathematics class amounted to $r = 0.78$ (Marjanovič Umek et al. 2006).

Moreover, Marjanovič Umek et al. (2006) determined that the factors that influence attainment (the characteristics of the learner and the environment) explain between 37% and 63% of the total variability in the learner's knowledge when the latter was evaluated with the National Assessments of Knowledge. Moreover, the same study determined that, between 56% and 62% of the total variability in

the learner's knowledge can be ascribed to the teacher-given grade. Researchers found that, despite pupils' attainments at the end of schooling differing very little from one another (i.e. they have similar grades in mathematics at the end of the school year), their achievements on the National Assessment of Knowledge differ significantly from the teacher-given grade. Systematically investigating the differences between the teacher-given grades and students' achievements on the National Assessment of Knowledge is, nevertheless, important, as it might offer educators and policymakers a better image of the evaluation standards. In particular, exploring the possible differences between the internal and external examinations of students' mathematical knowledge could give us better insights into grade inflation (Felda 2018), a phenomenon that occurs when the teacher-given grades are significantly higher than students' attainments on standardised tests.

Research

Purpose

By comparing the school years 2012/13, 2015/16, and 2018/19, this research examined whether the percentage of pupils with higher final grades is on the rise (cf. Felda 2018) and determined the correlation between pupils' score on the National Assessment of Knowledge and the final grade in mathematics at the end of primary school.

The final grades at the end of primary school were compared with the scores on the National Assessment of Knowledge in mathematics (which were converted into hypothetical grades) to determine whether the results indicate a potential subjectivity in the final mathematics grades.

Methodology

Non-experimental quantitative research method was applied.

Sample

Three samples of ninth grade Slovenian students were considered. The sample is representative since data is processed for the entire population of pupils at the end of primary school (Table 1) in all three school years and the National Assessment of Knowledge is compulsory for all pupils in Slovenia.

School year	2012/13	2015/16	2018/19
Number of pupils in the final year	17140	16648	16668

Table 1: Number of pupils in the population in a selected school year.

Data collection

The data, collected from the Slovenian National Examinations Centre after filing an official request, comprised the aggregated scores of pupils on the National Assessment of Knowledge in mathematics and the final grade in mathematics in the final year of primary school at the end of three school years: 2012/13, 2015/16, and 2018/19. Considering a larger time span (e.g. three evaluations in a period of six years) allows us to better understand the possible changes compared to a shorter period in time (e.g. three evaluations in consecutive years). The data was already anonymised; it was not possible to recognise individual students.

Data analysis

Both descriptive and inferential statistical methods were used. The study initially examined how many pupils with a specific score (expressed in percent) on the National Assessment of Knowledge in mathematics had the following final grades in mathematics: insufficient (1), sufficient (2), good (3), very good (4), or excellent (5). Descriptive statistics were used at this stage.

The method used to calculate the hypothetical grade was the same as the one that was used by the Mathematics Committee at the National Assessment of Knowledge in the school years from 2001/02 to 2004/05 and in Felda (2018). The committee abided by the distribution of grades given to pupils by mathematics teachers at the end of primary school (RIC 2005).

Let us give an example from 2012/13: out of 17140 pupils in the final year of primary school, 260 pupils (1.5%) were given the final grade »insufficient« (1). The same percentage of pupils (i.e. 1.5%) with the lowest scores on the National Assessment of Knowledge received the hypothetical grade »insufficient«. In reality, 206 pupils scored 12% of all points or less, while 301 pupils scored 14% of all points or less. That is why a correction was made. For the number of pupils with the final grade »insufficient« in Mathematics, i.e. 260, it is true that 206 < 260 < 301. A »less strict« version was chosen, and the hypothetical grade "insufficient" was assigned to pupils who scored 12% of all points or less. The lower limits were set for the other hypothetical grades similarly. The hypothetical grades were compared with the final grades obtained in mathematics class.

To sum up: the same percentage of pupils who were given »insufficient« as their final grade in school were also given the hypothetical grade »insufficient« on the National Assessment of Knowledge; the same percentage of pupils who were given the grade »sufficient« as their final grade in school were also given the hypothetical grade »sufficient« on the National Assessment of Knowledge. This procedure was conducted for each school year separately to determine the lower limit, expressed in percentage, for each hypothetical grade (Table 2).

Hypothetical grade	Lower limit in % for a specific hypothetical grade		
	2012/13	2015/16	2018/19
Sufficient (2)	14	8	12
Good (3)	44	38	36
Very good (4)	58	54	50
Excellent (5)	68	70	66

Table 2: Lower limits in determining the hypothetical grade on the National Assessment of Knowledge.

It has been established that the lower limits for the hypothetical grades »good« and »very good« have been lowering since the school year 2012/13 through 2015/16 to 2018/19; the difference in both hypothetical grades is 8 percentage points. The lower limits in the hypothetical grades »sufficient« and »excellent« vary; the difference between the highest and lowest lower limit in the hypothetical grade »sufficient« is 6 percentage points, and 4 percentage points in the hypothetical grade »excellent«. For the hypothetical grade »sufficient", the limit lowered in the 2015/16 school year compared to 2012/13 school year, and then rose in the 2018/19 school year. The opposite is true for the hypothetical grade »excellent«. In the 2015/16 school year, the limit rose compared to the 2012/13 school year and then lowered in the 2018/19 school year.

Pearson's correlation coefficient was used to calculate the correlation between the final grade in mathematics and the score on the National Assessment of Knowledge in mathematics, and between the pupil's final grade in mathematics and the hypothetical grade. A *t*-test was used to determine whether the difference between the final and hypothetical grades was statistically significant. The ANOVA test was used to check for differences in the means of the dependent variables among the three considered school years. To check the extent to which the teacher-given grades can predict pupils' hypothetical grades, a linear regression analysis was performed.

Results and discussion

Final grade in mathematics

The percentage of pupils with a specific final grade in the analysed school years (2012/13, 2015/16, and 2018/19) is roughly constant (Figure 1). The smallest deviation is noticeable in the final grade »insufficient«; there were 1.1% of pupils with this grade in the 2018/19 school year, and 1.8% in the 2015/16 school year, which makes up for a difference of 0.7 percentage points. The biggest difference is noticeable in the grade »excellent«, with 20.6% of pupils being given this grade in the 2012/13 school year, and as many as 22.3% in the 2018/19 school year.
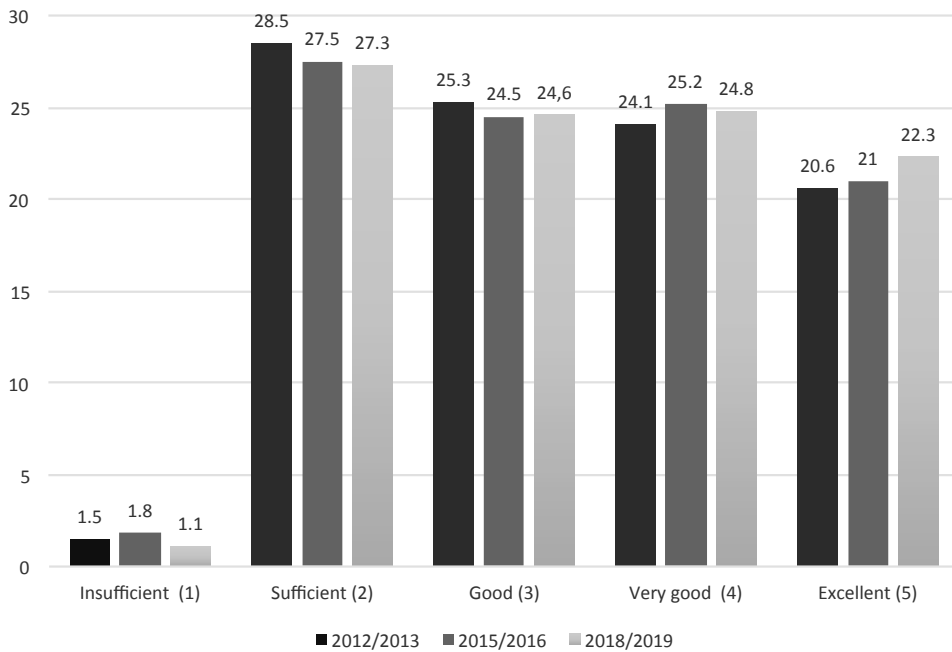
Figure 1: Percentage of pupils in the final year with a specific grade in mathematics

Table 3 presents the descriptive statistics of the final school grades in mathematics for the three school years. Despite the differences in means being relatively small, the ANOVA test revealed that the differences in the means of students' mathematics grades are statistically significant ($F(2;50453) = 12.1$; $p < .001$); however, the effect size is practically negligible ($\eta^2 = .000$).

School year	N	Mean	Standard deviation	Median	Minimum	Maximum
2012/13	17140	3.34	1.14	3	1	5
2015/16	16648	3.36	1.14	3	1	5
2018/19	16668	3.40	1.14	3	1	5

Table 3: The descriptive statistics of final school grades

Dispersion of scores on the National Assessment of Knowledge

Pupils with a specific final grade in mathematics achieved highly dispersed scores on the National Assessment of Knowledge in mathematics. Pupils with the final grade »sufficient« scored from 0% to around 80% of all points; those with the final grade »good« scored from 0% to around 94% of all points, and those with the final grade »very good« from 0% to around 96% of all points. Pupils with the final

grade »excellent« scored from 14% to 100% (Figure 2). It has been determined that only pupils with the final grade »excellent« scored all the points on the National Assessment of Knowledge in mathematics, whereas a score of 0% was observed in pupils with final grades ranging from »insufficient« to »very good«. A thorough review of the data (Figure 2) revealed that the highest percentage of points scored on the National Assessment of Knowledge in mathematics with regard to each final grade has not deviated considerably over the years. The biggest deviation (by 8 percentage points) was noticeable in pupils with the final grade »excellent« :in the school years 2012/13 and 2015/16, the lowest score attained by such pupils was 14%; in the 2018/19 school year it amounted to 22%.
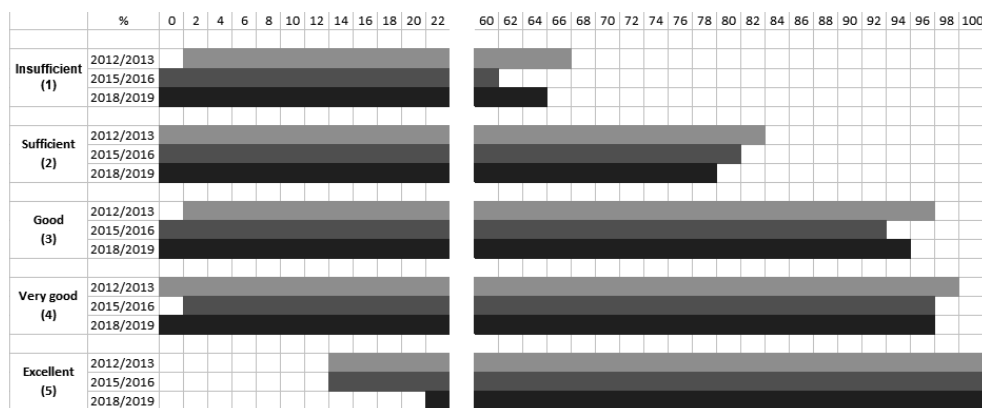

Figure 2: Dispersion of scores on the National Assessment of Knowledge in mathematics with regard to the final grade in mathematics.

A more thorough review of the data also points out »loners« (e.g. in Table 4 for the 2012/13 school year). They are those rare pupils with the only lowest score; pupils with the next high score appear much later (e.g. in the 2012/13 school year, one pupil with the final grade »very good« scored 0% on the National Assessment of Knowledge in mathematics; the next two pupils scored 14%, and were then continuously followed by pupils with scores of 20% and higher).

Final grade	Score on the National Assessment of Knowledge (in %)													
	0	2	4	6	8	10	12	14	16	18	20	22	24	26
Insufficient (1)	0	3	1	3	3	9	12	14	8	12	13	14	11	19
Sufficient (2)	3	8	7	16	21	46	63	76	103	121	172	166	195	203
Good (3)	0	1	0	0	3	4	2	2	9	9	25	23	36	36
Very good (4)	1	0	0	0	0	0	0	2	0	0	2	2	1	3
Excellent (5)	0	0	0	0	0	0	0	1	0	0	1	0	0	1

Table 4: Showing the »loners« in the 2012/13 school year

Average scores on the National Assessment of Knowledge

The average score or the average number of points in percent achieved on the National Assessment of Knowledge in mathematics is always around 50%; the exact data for each school year is shown in Table 5. The ANOVA test revealed that, among the three considered school years, there are statistically significant differences ($F(2;50453) = 204$; $p < .001$) with a small effect size ($\eta^2 = .008$).

School year	N	Mean	Standard deviation	Median	Minimum	Maximum
2012/13	17140	55.3	19.8	56	0	100
2015/16	16648	51.5	21.8	52	0	100
2018/19	16668	51.1	21.5	50	0	100

Table 5: The descriptive statistics for the National Assessment of Knowledge in Mathematics (all results are expressed in percentage points).

The average value of scores on the National Assessment of Knowledge in mathematics with regard to the final grade in mathematics was calculated using the weighted arithmetic mean. It was then determined that the average values in the school years 2015/16 and 2018/19 are very similar; a smaller deviation is noticeable in 2012/13 (Table 6). It has been established that the deviation in the 2012/13 school year, when compared to the 2015/16 and 2018/19 school years, is biggest in the case of the grade »insufficient« (by 7.3 percentage points) and smallest in the case of the grade »excellent« (by 2.7 percentage points).

	Average score in %		
Final grade	2012/13	2015/16	2018/19
Insufficient (1)	28.8	21.8	21.5
Sufficient (2)	37.0	31.7	31.2
Good (3)	50.9	46.4	44.6
Very good (4)	63.8	60.4	58.8
Excellent (5)	78.0	75.3	75.4

Table 6: Average value of scores on the National Assessment of Knowledge in Mathematics with regard to the final grade in mathematics.

Hypothetical grades

Table 7 presents the descriptive statistics for the obtained hypothetical grades. The means of the hypothetical grades in different school years differ significantly ($F(2;50453) = 22.8$; $p < .001$), while the effect size is almost negligible ($\eta^2 = .001$).

School year	N	Mean	Standard deviation	Median	Minimum	Maximum
2012/13	17140	3.47	1.20	3	1	5
2015/16	16648	3.43	1.15	3	1	5
2018/19	16668	3.52	1.18	4	1	5

Table 7: Descriptive statistics for hypothetical grades

The results, presented in tables 8 to 10, show the dispersion of National Assessment of Knowledge scores and the individual final grades. It has been established that none of the pupils with the final grade »insufficient« would have been given the hypothetical grade »excellent«, and that none of the pupils with the final grade »excellent« would have been given the hypothetical grade »insufficient«.

Around 1% of the pupils with the final grade »insufficient« would have been given the hypothetical grade »very good«. In recent years, the percentage of pupils who would have been given the hypothetical grade »good« has been in decline (from 15.38% in the 2012/13 school year to 5.88% in the 2018/19 school year), whereas the percentage of pupils with the hypothetical grade »sufficient« has been on the rise (from 71.54% in 2013 to 82.35% in the 2018/19 school year). The percentage of pupils with the hypothetical grade »insufficient« varies from 11.92% in the 2018/19 school year to 14.43% in the 2015/16 school year and to 10.70% in the 2018/19 school year. The percentage of pupils with the final grade »insufficient« who would have been given the hypothetical grade »sufficient« increased in the examined years by more than 10 percentage points; at the same time, the percentage of pupils who would have been given the hypothetical grade »good« decreased by almost as much. The question arises whether teachers demand more than they should for the final grade »sufficient« and unjustifiably give pupils the grade »insufficient«, making them resit the exam.

The percentage of pupils among those with the final grade »sufficient« who would have been given the hypothetical grade »excellent« is less than 1, while the percentage of pupils who would have been given the hypothetical grade »very good« ranges from 5.46% to 8.49%. The percentage of pupils who would have been given the hypothetical grade »good« is in a slight decline (from 28.38% in the 2012/13 school year to 25.74% in the 2018/19 school year); the percentage of pupils who would have been given the hypothetical grade »sufficient« is rather constant (from 59.64% to 61.88%), as is the percentage of pupils who would have been given the hypothetical grade »insufficient« (from 3.30% to 3.93%).

The percentage of pupils among those with the final grade »good« who would have been given the hypothetical grade »excellent« ranges from 6.06% in the 2015/16 school year to 9.48% in the 2012/13 school year. There is a moderate increase in the percentage of pupils who would have been given the hypothetical grade »very good«, namely from 23.30% in the 2012/13 school year to 28.90% in the 2018/19 school year. There is a slight decline in the percentage of pupils with the hypothetical grade »good« (from 40.93% in the 2012/13 school year to 36.94% in the 2018/19 school year). The percentage of pupils with the hypothetical grade

»sufficient« is constant (around 26%). There is less than half a percent of pupils with the final grade »good« and the hypothetical grade »insufficient«.

The percentage of pupils among those with the final grade »very good« who would have been given the hypothetical grade »excellent« or »very good« varies greatly. The difference in the hypothetical grade »excellent« is 12.93 percentage points, and 11.72 percentage points in the grade »very good«. The percentage of pupils with the hypothetical grade »good« is rather constant (between 20.15% and 22.35%), as is the percentage of pupils with the hypothetical grade »sufficient« (between 4.72% and 6.36%). The percentage of pupils with the hypothetical grade »insufficient« is very low (between 0.02% and 0.15%).

The percentage of pupils with the hypothetical grade »excellent« varies considerably (from 81.01% in the 2012/13 school year to 69.89% in the 2015/16 school year and to 77.59% in the 2018/19 school year). The percentage of pupils who would have been given the hypothetical grade »very good« also varies (from 13.99% in the 2012/13 school year to 23.48% in the 2015/16 school year and to 17.94% in the 2018/19 school year). A smaller deviation is noticeable in the percentage of pupils who would have been given the hypothetical grade »good« (less than 2 percentage points). There is less than one percent of pupils with the final grade »excellent« and the hypothetical grade »sufficient«.

It can be seen that the percentage of pupils with the final grade »very good« who would have been given the hypothetical grade »excellent« dropped by more than 10 percentage points between the school years 2012/13 and 2015/16; it then increased by more than 6 percentage points in the 2018/19 school year. The contrary is true for the percentage of pupils with the final grade »excellent« who would have been given the hypothetical grade »very good«, which increased by just under 10 percentage points between the school years 2012/13 and 2015/16 and then dropped by more than 6 percentage points in the 2018/19 school year.

Final grade at the end of schooling	Hypothetical grade on the National Assessment of Knowledge				
	Insufficient (1)	Sufficient (2)	Good (3)	Very good (4)	Excellent (5)
Insufficient (1)	11.92	71.54	15.38	1.15	.00
Sufficient (2)	3.36	61.88	28.38	5.46	.92
Good (3)	.23	26.05	40.93	23.30	9.48
Very good (4)	.02	4.81	22.69	31.26	41.22
Excellent (5)	.00	.40	4.60	13.99	81.01

Table 8: The percentage of pupils with the hypothetical grade in relation to a specific final grade in mathematics at the end of schooling in the school year 2012/13.

Final grade at the end of schooling	Hypothetical grade on the National Assessment of Knowledge				
	Insufficient (1)	Sufficient (2)	Good (3)	Very good (4)	Excellent (5)
Insufficient (1)	14.43	73.49	11.41	.67	.00
Sufficient (2)	3.39	59.64	28.17	7.60	.66
Good (3)	.34	25.58	40.38	27.64	6.06
Very good (4)	.02	6.36	22.35	42.98	28.29
Excellent (5)	.00	.83	5.80	23.48	69.89

Table 9: The percentage of pupils with the hypothetical grade in relation to a specific final grade in mathematics at the end of schooling in the school year 2015/16

Final grade at the end of schooling	Hypothetical grade on the National Assessment of Knowledge				
	Insufficient (1)	Sufficient (2)	Good (3)	Very good (4)	Excellent (5)
Insufficient (1)	10.70	82.35	5.88	1.07	.00
Sufficient (2)	3.30	61.86	25.74	8.49	.62
Good (3)	.17	26.07	36.94	28.90	7.92
Very good (4)	.15	4.72	20.15	40.25	34.73
Excellent (5)	.00	.59	3.88	17.94	77.59

Table 10: The percentage of pupils with the hypothetical grade in relation to a specific final grade in mathematics at the end of schooling in the school year 2018/19

Correlation between students' achievements

This study examined the correlation between pupils' grade in mathematics and their score on the National Assessment of Knowledge in mathematics. The survey encompassed all the students who took the National Assessment of Knowledge. For the school year 2012/13 ($r = .770$; $p < .001$), 2015/16 ($r = .751$; $p < .001$), and 2018/19 ($r = .770$; $p < .001$), the Pearson's correlation coefficient indicated a positive and rather high correlation, indicating that pupils with higher final grades in mathematics also scored higher on the National Assessments of Knowledge in mathematics.

Each pupil was assigned a hypothetical grade. The hypothetical grade average (Table 7) ranges from 3.43 to 3.52. Table 11 shows the average differences between final grades and hypothetical grades. The average difference between the final grades and the hypothetical ones varies from -0.0682 to -0.137. With a *t*-test for independent samples, it can be shown that hypothetical grades are statistically higher than the average final grades ($p < .001$), as anticipated, because the lower limits for the hypothetical grades were defined based on the final grades but under less strict conditions, which means that the lower bounds might be estimated less strictly as the one presented in the Data analysis section.

	School year		
	2012/13	2015/16	2019/19
Average difference between the final and the hypothetical grades	-0.137 $SE = 0.006$	-0.0682 $SE = 0.007$	-0.119 $SE = 0.006$
t-test	$t(17139) = -21.9$ $p < 0.001$	$t(16647) = -10.5$ $p < 0.001$	$t(16667) = -18.7$ $p < 0.001$

Table 11: The differences between final grades and hypothetical grade average, and the results of the t-test

Lastly, the study examined how many pupils would have been given a hypothetical grade higher than the final grade, how many would have been given the same grade, and how many a lower grade than the final one. It has been determined that just over half of the pupils would have been given a hypothetical grade that equals the final grade; roughly one-fifth of the pupils would have been given a hypothetical grade lower than the final grade; and just under a third of the pupils would have been given a hypothetical grade higher than the final grade (Figure 3).
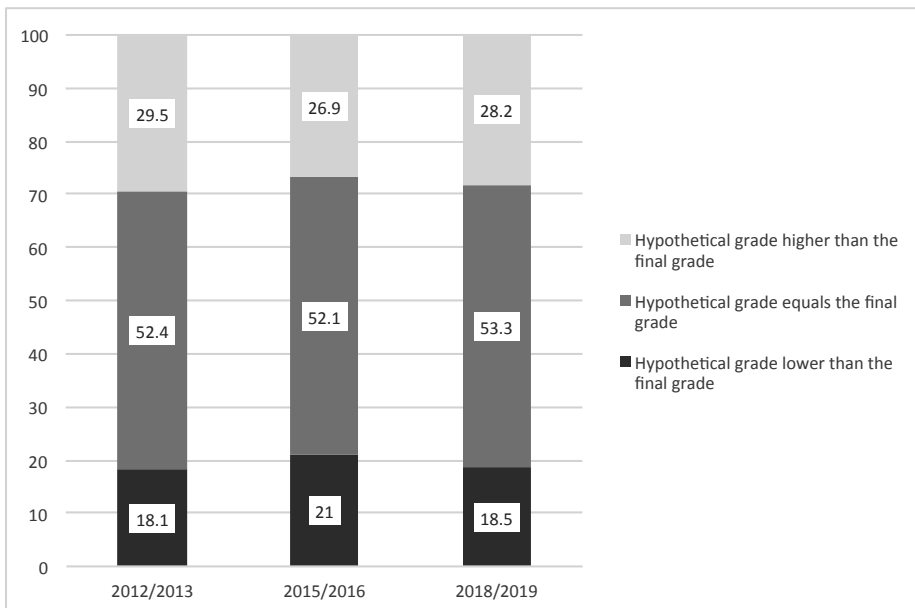

Figure 3: Percentage of pupils with specific hypothetical grades with regard to the final grade

After converting the scores (expressed in percentage) on the National Assessment of Knowledge in mathematics into hypothetical grades that the pupils would have been given if only these scores were taken into account, it was examined whether there was a correlation between these scores and the pupils' final grades. The Pearson's correlation coefficient indicates a rather high positive and statistically significant correlation between the two variables (Table 12). In this case too, it can be said that the pupils with higher school grades would have been given higher hypothetical grades.

	School year		
	2012/13	2015/16	2018/19
Pearson's correlation coefficient between hypothetical grade and final grade	$r = 0.753$ $p < 0.001$	$r = 0.731$ $p < 0.001$	$r = 0.750$ $p < 0.001$

Table 12: Correlation between hypothetical grade and final grade

To verify the extent to which the hypothetical grades were predicted by the teacher-given grades, linear regression was conducted. In the linear model, the hypothetical grades were treated as dependent variable, the teacher-given grades as independent variables, and the school years as parameters. The model was significant ($F(3;50452) = 20992$; $p < .001$), and it explained 55.5% of the variance. The teacher-given grade was a significant predictor ($\beta = .745$; $t = 250.75$; $p < .001$), while the only significant difference between school years was the one between 2015/16 and 2012/13 ($\beta = -.0540$; $t = -7.45$; $p < .001$); the difference between the school year 2018/19 and 2012/13 was not significant ($\beta = -.00374$; $t = -.516$; $p = .606$).

Conclusions

Teachers determine the final grade in mathematics, which is entered in their school-leaving certificate, at the end of the school year based on the grades awarded to pupils throughout the school year. This study determined that, over the years, the percentage of pupils with the final grade »excellent« was on the rise, the percentage of pupils with the grade »sufficient« was in decline, and the percentages of pupils with the other final grades varied. The individual learning differences in the pupils' academic performance can be explained by their speech competence, their intellectual ability, their personality dimensions (conscientiousness and openness/intellect), and the differences in some of the components of parental pressure on school work and in the education level of the pupils' parents (Marjanovič Umek et al. 2006).

For the school years 2015/16 and 2018/19, the average score on the National Assessment of Knowledge in mathematics in relation to the final grade is constant in the case of the final grades »excellent«, »sufficient« and »insufficient«; a smaller deviation is noticeable in the final grades »good« and »very good«. The average score on the National Assessment of Knowledge of pupils with the final grades »excellent« and »very good« was found to be higher than the state average, whereas pupils with final grades »good«, »sufficient« and »insufficient« were found to score below the state average on the National Assessment of Knowledge. It can be concluded that there exists a percentage of pupils who would receive a higher final grade than the one they were given and vice versa—that is, there is a percentage of pupils who deserve a lower final grade.

This study confirms that there is a high correlation between the hypotheti-

cal grade obtained at the National Assessment of Knowledge and the final grade in mathematics at the final year of primary school. This is demonstrated by the statistically significant correlation coefficients between the score on the National Assessment of Knowledge and the final grade, and between the final grade and the hypothetical grade. This correlation varies and does not indicate a constant trend (Table 8 and Table 10). This is evident from the dispersion of hypothetical grades within the final grades.

The correlation between pupils' grades in mathematics and the score on the National Assessment of Knowledge in mathematics indicates a positive, rather high, and statistically significant correlation. It can be concluded that pupils with higher final grades in mathematics also score higher on the National Assessments of Knowledge, which is also indicated by the calculated percentage of pupils with a specific hypothetical grade and final grade in mathematics. The t-test determined that the hypothetical grades are statistically higher than those given to the pupils in their school-leaving certificates. Moreover, in just under half of the pupils, the hypothetical grade was found to differ from the final grade (a higher or lower grade). This percentage is high and is not dropping, meaning that just over half of the pupils have the same final and hypothetical grade. It can be deduced that certain pupils were given an unfair final grade, and that this percentage is not negligible.

Lekholm and Cliffordson (2008) claimed teachers' orientation to be one reason for many pupils getting better final grades, especially on account of the negative effects a bad final grade would have on an individual pupil and on the school (e.g. repeating a year). They added that competition among schools to enrol pupils (in Slovenia, primary schools are defined by school districts) makes teachers award higher final grades to compensate for pupils' poor National Assessments of Knowledge scores. Therefore, such practices lead to differences in grades and differences between schools. This has also been confirmed by other studies (Cross and Frary 1999; Wikström 2005).

This study was also interested in the correlation between the final and the hypothetical grade. The Pearson's correlation coefficient indicates a rather high positive and statistically significant correlation, indicating that pupils with higher final grades have higher hypothetical grades; however, there exists no perfect linear match between the two variables. Even though pupils' final grades were found to match quite well with their scores on the National Assessments of Knowledge, the correlation is not perfect. Various other factors also influence pupils' scores on the National Assessments of Knowledge and their final grades. A vast majority of pupils with the final grade »excellent« would have been given the hypothetical grade »excellent«, while not even half of the pupils with the final grade »very good« would have been given the hypothetical grade »very good« (Tables 8–10). Likewise, the proportion of pupils with the final grade »good« who would have been given the hypothetical grade »good« is less than half, while roughly a quarter of such pupils would have been given the grade »very good« and roughly a quarter the grade »sufficient« (Tables 8–10). Around 60% of the pupils with the final grade »sufficient« would have been given the hypothetical grade »sufficient«, and more

than 70% of the pupils with the final grade »insufficient« would have been given the hypothetical grade »sufficient« (Tables 8–10).

When explaining the comparisons between the final grade and the score on the National Assessment of Knowledge, one must take into account that the final grade and the National Assessment of Knowledge score or the hypothetical grade are two different criteria of academic performance. The differences between the final grade and the hypothetical grade can be explained by the fact that teachers test the knowledge and skills acquired in class in different ways (through, for example, written or oral tests), whereas the National Assessment of Knowledge contains only a written test. The final grade is the result of the pupils' previously obtained grades (at least six per school year), while the National Assessment of Knowledge is a one-time event that covers the learning contents of three school years. The National Assessment of Knowledge does not assess all the target knowledge (it does not, for example, contain tasks that assess the use of a calculator). Other causes lie in the motivation and effort invested by pupils in taking the test.

Hence, we should keep fairness in mind even when giving final grades, because a fair system should strive towards making the results of education and training independent of socioeconomic background and other factors, which lead to educational disadvantage (Commission of the European Communities 2006). Hauptman (2012) determined that the only thing that will change teachers' behaviour and, consequently, change pupils' attainments is a change in teachers' opinions on self-evaluation. However, this takes longer to realise.

This research is not without limitations. Since national assessment grades are hypothetical and lower margins should be selected when setting boundaries for them, this can cause a greater percentage of high grades. These lower limitations are not constant in time; therefore, future studies can examine different methods to convert achievements on national examinations into school grades.

References

Cankar, G., Brelih-Hatunič, A., Grašič Arnuš, I., Gorše, A., Jerše, L., Ozimek, B., Rupnik Vec, T., Špilak, B. and Zupanc, B. (2019). *Dosežki učencev ter dosežki otrok v razvoju in učenju*. Ljubljana: Šola za ravnatelje.

Commission of the European Communities. (2006). *Communication from the Commission to the Council and to the European Parliament: Efficiency and equity in European education and training systems.* Retrieved from: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52006DC0481 (assessed on 8. 6. 2022).

Cross, L. H. and Frary, R. B. (1999). Hodgepodge grading: Endorsed by pupils and teachers alike. *Applied measurement in Education,* 12, issue 1, pp. 53–72.

Felda, D. (2018). Preverjanje matematičnega znanja. *Journal of Elementary Education,* 11, issue 2, pp. 175–188.

Hauptman, A. (2012). Spreminjanje prepričanj učiteljev o samoevalvaciji šole ter vpliv na vedenje učiteljev in dosežke učencev. *Psihološka obzorja,* 21, issue 2, pp. 19–28.

Kartianom, K. and Mardapi, D. (2017). The utilization of junior high school mathematics national examination data: A conceptual error diagnosis. *Research and Evaluation in*

Education, 3, issue 2, pp. 163–173.

Khadijeh, B. and Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education,* 3, issue 1, pp. 139–146.

Lekholm, A. K. and Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation,* 14, issue 2, pp. 181–199.

Lindahl, E. (2007). C*omparing teachers' assessments and national test results: evidence from Sweden* (No. 2007: 24). Working Paper.

Marjanovič Umek, L., Sočan, G. and Grgić, K. (2006). Šolska ocena: koliko jo lahko pojasnimo z individualnimi značilnostmi mladostnika in koliko z dejavniki družinskega okolja. *Psihološka obzorja,* 4, issue 15, pp. 25–52.

Metsämuuronen, J. and Ukkola, A. (2022). Rudimentary stages of the mathematical thinking and proficiency: Mathematical skills of low-performing pupils at the beginning of the first grade. *LUMAT: International Journal on Math, Science and Technology Education,* 10, issue 2, pp. 56–83.

Organisation for Economic Co-operation and Development. (2008). *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools. OECD Publishing and Centre for Educational Research and Innovation.* Retrieved from https://www.oecd-ilibrary.org/education/measuring-improvements-in-learning-outcomes_9789264050259-en (Assessed on 6. 8. 2020)

RIC. (2005). *Nacionalni preizkusi znanja. Letno poročilo o izvedbi v šolskem letu 2004/2005. Ljubljana: Državni izpitni center.* Retrieved from https://www.ric.si/mma/letno%20 poro%C4%8Dilo%20npz%202005/2006070311492832/ (assessed on 7. 8. 2020).

Smole, D. (1999). Korelacija zaključne ocene pri matematiki in slovenščini v 8. razredu z uspehom pri zunanjem (eksternem) preverjanju v šolskem letu 1996/1997. *Didakta,* 8, issue 44/45, p. 76.

Wikström, C. (2005). *Criterion-referenced measurement for educational evaluation and selection* (Doctoral dissertation, Beteendevetskapliga mätningar, Umeå universitet).

Wößmann, L. (2008). Efficiency and equity of European education and training policies. *International Tax and Public Finance,* 15, issue 2, pp. 199–230.

Wyatt-Smith, C., Klenowski, V. and Colbert, P. (Eds.). (2014). *Designing assessment for quality learning (Vol. 1).* Springer Science & Business Media.

Jerneja BONE (Zavod RS za šolstvo, Slovenija)
Daniel DOZ (Univerza na Primorskem, Pedagoška fakulteta, Slovenija)

PRIMERJAVA DOSEŽKOV UČENCEV MED NACIONALNIMI PREIZKUSI ZNANJA IN ZAKLJUČNIMI OCENAMI IZ MATEMATIKE

Povzetek: Smiselna analiza podatkov, ki jih pridobimo na internih in eksternih preverjanjih znanja, je pomembna za načrtovanje nadaljnjega poučevanja. V prispevku prikažemo longitudinalno primerjavo dosežkov učencev 9. razreda na nacionalnih preizkusih znanja matematike in zaključnih ocen ob zaključku šolskega leta. Primerjali smo podatke v treh obdobjih, z razmikom treh let. Ugotavljali smo, ali se viša odstotek učencev z višjimi zaključnimi ocenami in ali je povezanost med hipotetično oceno pridobljeno na Nacionalnem preverjanju znanja in zaključno oceno pri matematiki v 9. razredu osnovne šole visoka. V raziskavi smo ugotovili, da je odstotek učencev z zaključno oceno odlično v rahlem porastu ter da je odstotek učencev z oceno zadostno v rahlem upadu. Potrdili smo, da obstaja visoka povezanost med hipotetično oceno pridobljeno na nacionalnem preverjanju znanja in zaključno oceno pri matematiki v 9. razredu osnovne šole. V raziskavi smo upoštevali, da sta zaključna ocena in dosežek na nacionalnem preverjanju znanja oz. hipotetična ocena dve različni merili šolske uspešnosti.

Ključne besede: eksterno preverjanje znanja; interno preverjanje znanja; hipotetična ocena; matematično znanje; ocenjevanje

Elektronski naslov: jerneja.bone@zrss.si