Daniel Doz, M.Sc., University of Primorska, Faculty of Education, Koper, SLOVENIA, National Scientific Lyceum »France Prešeren« with Slovene as language of instruction, Trieste, ITALY;

Darjo Felda, University of Primorska, Faculty of Education,

**GRADE INFLATION IN ITALIAN HIGH SCHOOLS: EVIDENCE FROM THE NATIONAL ASSESSMENT INVALSI**

**INFLACIJA OCEN NA ITALIJANSKIH SREDNJIH ŠOLAH: DOKAZ IZ NACIONALNEGA PREIZKUSA ZNANJA INVALSI**

# **Introduction**

Grading is a key element of every pedagogical process, since one of its primary aims is to understand students’ real competences and knowledge (Kartianom & Mardapi, 2017; Pramudita et al., 2019), while giving teachers important feedback information about their students’ understanding (Olutola et al., 2019; Felda, 2018).

However, when it comes to national assessments, these rarely have formative functions, since they are mostly made with diagnostic purposes, that is, their aim is to assess the schooling system of a specific country (Felda, 2018; Bansilal, 2017). This would be difficult to do if one would merely rely on teachers’ grades (Bauer & Sheldon, 2008), because these are subjective and they do not only consider students’ knowledge, but also various conative factors (Kunnath, 2017; Özdemir & Özkan, 2017; Rauschenberg, 2014). Subjective grading could lead to the issue of grade inflation, i.e. when teachers’ grades are higher than the real knowledge level of individual students; hence students, who would need additional help, would get an unrealistic feedback about their real competences (Gershenson, 2018).

Although teachers’ grades might differ from students’ achievements on standardized tests, school grades serve many purposes, for instance they provide information about students’ attitude, hard-working, writing homework etc. (Feron et al., 2016). However, even students that have excellent grades in school fail to show the same performances in national assessments (Gershenson, 2018; Bauer & Sheldon, 2008). Teachers’ grades are more subjective; however, they include more information about students’ attitudes. National assessments, on the contrary, do not contain as much information as the teachers’ grades, but they are a more objective way of assessing students’ knowledge (Nordin et al., 2019; Lindahl, 2007). It has been proved that students’ school grades and achievements on the national assessments are correlated (Gershenson, 2018), however some differences among them have been found.

Hence, the results of national assessments should be considered as complementary to teachers’ grades: one result should not exclude the other (Gershenson, 2018; Finefter-Rosenbluh & Levinson, 2015), since with both results, we could get a clearer image of the students’ knowledge and competences. Felda (2018) affirms that considering the results of national assessments as part of the final students’ grade could help to deflate excessively high school grades, giving a more objective image about students’ knowledge and could contribute to a more normal distribution of grades. In the literature, however, no standardized methods of assessing students’ mathematical knowledge were proposed, hence there are no clear indications about how legislators and educators should consider both grades to have a more objective picture about students’ knowledge.

Generally, research in the field of grade inflation investigated the differences between school grades and performances on standardized tests, concluding that there is evidence of this issue, whether bigger discrepancies are present between the two grades (Felda, 2018). In addition, several researches proved the existence of grade inflation by analyzing students’ average grades in a specific time period. However, in both cases, individual differences between school grades and national assessments were rarely considered, since the focus of such researches was to identify evidence of grade inflation in bigger cohorts and in longer time periods.

In the present work, we aim to contribute to the literature on the topic of grade inflation by examining more in detail such phenomenon in Italian high schools. To this day, no systematic research was made to verify the presence of grade inflation on the Italian territory. Such research question is nevertheless important, since, as highlighted in some of the previously mentioned works, grade inflation could lead legislators, principals, educators and students to have a wrong picture about students’ knowledge and the overall quality of the national educational system. Moreover, being the measurement of mathematics knowledge particularly important, since mathematical skills and competencies are used in everyday life (Geary et al., 2013), having an objective picture about students’ knowledge is particularly important.

Guided by the idea that grade inflation is a problem of global concern (Nordin et al., 2019), the main goal of this paper is to provide empirical evidence on the presence of grade inflation in mathematics in Italian high schools. More specifically, we were interested in investigating whether and to what extent do teachers’ mathematics grades differ from students’ achievements on national assessments. Furthermore, we analyze a way of converting students’ achievements on the Italian national assessment of mathematics INVALSI into school grades, and we use these to make additional considerations about how educators and policymakers could use them, in order to have a clearer picture of students’ mathematical knowledge.

# **Theoretical Framework**

## *Grade inflation*

Understanding teachers’ grading strategies is a difficult issue, since we have to consider not only the objective evaluation of students’ knowledge and competences, but also the affective and subjective interferences, which play a significant part in any grading process. The willingness of helping students to access a prestigious university or the need of awarding students’ motivation and their constant hard work (Kunnath, 2017; Özdemir & Özkan, 2017; Keller, 2016; Rauschenberg, 2014), as well as gender, racial and other prejudices (Rauschenberg, 2014) are among the factors that can interfere with a more objective grading system. What these interferences ultimately leads us to, is grade inflation.

Inflation of school grades occurs whenever teachers’ grades are higher than the objectively measured students’ competences and abilities. What this means is that students’ grades, rather than reflecting their real knowledge, are a reflection of a teachers’ perception of what that knowledge means and therefore are put on higher points in the grading scale, while students’ knowledge remains constant or does even decrease (Nordin et al., 2019; Chowdhury 2018; Gershenson 2018; Caruth & Caruth 2013). Since grading scales are upward limited, grade compression might occur (Nordin et al., 2019; Chowdhury 2018; Gershenson 2018; Butcher et al., 2014), so that a higher density of occurrences is seen in the higher part of the grading.

The issue is that grades inflation can lead to a confusing feedback about students’ achievements. This is because it shows an unrealistic picture of students’ knowledge and understanding which is ultimately detrimental to the students themselves who believe they mastered their coursework when in reality they have not (Chowdhury, 2018; Finefter-Rosenbluh & Levinson, 2015; Butcher et al., 2014). Hence, addressing the issue of grade inflation is an important research question, since unreal and inflated grades could lead to a wrong picture of the national educational system. Such distorted picture might lead educators and policymakers not to act immediately to provide extra help to underachieving students.

Unfortunately, the inflation of school grades is not an isolated problem, but a worldwide phenomenon, which has been studied by many in the last decades (Nordin et al., 2019; Chowdhury, 2018; Ali et al., 2016; Finefter-Rosenbluh & Levinson, 2015; Costley, 2014). For example, in the U.S.A. the percentage of excellent students rose from 39% to 47% from the year 1998 to 2016, at the same time, however, we can see a decrease in the results of SAT tests (Gershenson, 2018; Hurwitz & Lee, 2018). In Sweden, the average grades rose about 10% from 1997 to 2003, while the results of Swedish students on international standardized tests (e.g. PISA) remained mainly constant (Nordin et al., 2019). Similarly, in Pakistan the average grade rose, while students’ competences and knowledge decreased (Khan & Hussain, 2019; Ali et al., 2016; Ayubbuzder & Ali, 2013). Such considerations are valid also for Slovenia: comparing students’ grades with their results in the national assessment *Nacionalno Preverjanje Znanja* (NPZ) we can conclude that only 69.89% of excellent students had an excellent performance in the NPZ (Felda, 2018). In such research, grade inflation was detected by converting students’ achievements on national assessments into school grades. The comparison between the number of students who got an excellent grade and those, who had an excellent performance on the national assessment, showed that school grades tend to be higher than those students would get on national assessments.

In the international literature, grade inflation was studied mainly from two perspectives. Firstly, some works showed that, in a more or less long time period, average students’ grades increased, hence proving that teachers grade more leniently (Bar et al., 2009; Babcock, 2010; Bachan, 2017; Butcher et al., 2014; Jewell et al., 2013). In such kind of studies, authors analyzed the trends in average students’ grades (GPAs) and degrees in years, noticing that both grade compression and grade inflation is occurring. In particular, in the last years students’ grades rose. Evidence of grade inflation is, therefore, believed to be associated to increasing grades in time.

Secondly, some researches focused on comparing students’ average grades (GPAs) with their achievements on standardized tests, such as national and international assessments (Ziomek & Svec, 1997; Wikström & Wikström, 2005; Pattinson et al., 2013). Results suggest that students tend to have higher school grades than results on standardized assessments of knowledge. Moreover, when mean grades and achievements on standardized tests were considered in a time period, school grades tend to increase, while results on national tests are almost constant in time. Therefore, teachers give higher grades for the same knowledge.

Students’ achievements on national assessments are not always expressed in terms of school grades. Hence, the necessity of converting achievements into school grades represents an additional issue while studying the phenomenon of grade inflation. For instance, Felda (2018) converted students’ achievements on the Slovenian national assessment from percentages into school grades (from 1 to 5), analyzing then the differences between such converted grades (hereby, hypothetical grades) with actual teachers’ grades. The author suggested to develop a way of assessing students’ mathematical knowledge by considering both hypothetical grades, which are the results of the conversion of achievements on national assessments into school grades, and teachers’ grades, confirming the concerns of some other researchers (Gershenson, 2018; Finefter-Rosenbluh & Levinson, 2015).

## *Teachers’ grades and national assessments*

Teachers’ grades serve many purposes, in fact, give information about a student’s attitude toward the subject, how hard did he/she work, whether he/she was regular in turning in homework etc. (Feron et al., 2016). However, some researchers have shown that students that have excellent grades in school fail to show the same performances while solving some problems from the national assessments (Gershenson, 2018; Bauer & Sheldon, 2008). Hence, the question about how to grade students’ knowledge of mathematics emerges. From one point of view, teachers’ grades are more subjective, but they include more information about students’ achievements, class participation, homework etc. National assessments do not contain as much information as the teachers’ grades, but they are a more objective way of assessing students’ knowledge, lowering the probability that various discriminations and prejudices would influence the results, while grade inflation is less present (Nordin et al., 2019; Lindahl 2007). Even if grades and results in standardized tests are correlated (Gershenson 2018; Brennan et al., 2001; Willingham et al., 2002; Bowers, 2011; Argentin & Triventi, 2015), they might considerably differ.

The researches made in the United States indicated a middle-strong correlation between students’ achievements on different types of standardized assessments and their school grades (Brennan et al., 2001; Willingham et al., 2002; Bowers, 2011). For what it concerns Italy, less research was made to study this topic. Argentin and Triventi (2015) found a correlation of *r=*.33 between grade 10 students’ mathematics achievements on the INVALSI test and school grades.

## *The National Assessment of Mathematics in Italy*

The National Evaluation System (*Sistema Nazionale di Valutazione,* SNV) works as part of the National Institute for the Evaluation of the schooling and instruction system (*Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione*, INVALSI). The main aim of the SNV is to periodically assess the quality of the Italian schooling and instruction system by testing students’ knowledge of the Italian (or other minority) language, mathematics and English. (Legislative Decree n. 286/2004, Ministry Decree MIUR 76/2009). Students are required to take the assessment of knowledge in grade 2, 5, 8, 10 and 13 (Quadro di Riferimento, 2018).

The INVALSI Institute does not grade students with standardized grades that are present in the Italian evaluation system (i.e. increasing grades from 1 to 10, where 6 is the first positive grade). Results on the INVALSI assessments are computed with a special metric which is calculated through a selection/sample of students who take the test a month prior to all other schools. The results are then converted into the INVALSI metric (INVALSI, 2018). Trough the analysis of various psychometric quantities, where the Institute defines the origin (the “zero”) of the scale, while distributing the results linearly on the scale, considering 200 as the mean of the results, with a standard deviation of 40.

Based on this metric, the INVALSI considers five different levels of knowledge. For example, here we present the levels of knowledge for grade 10 students (Rapporto, 2019):

* 1st level: the student uses only an elementary knowledge, which he/she has gained in the previous schooling levels; the student answers to elementary questions about topics that he/she has already encountered in class;
* 2nd level: the student does understand elementary concepts from the national program form the first two years of high school and he/she knows the basic computing strategies, uses elementary algorithms and solves problem using merely elementary procedures;
* 3rd level: the student uses his/her knowledge of the first two years of high school and does connect different element, he/she answers to questions that require elementary reasoning, recognizing the different mathematical elements and properties;
* 4th level: the student recognizes, also in unusual situations, mathematical objects (such as equations), and know to use them to solve more complex mathematical problems. He/she interprets correctly the given data and connects different mathematical fields. The student is able to use the formal mathematical and symbolic language and is capable of mathematical reasoning;
* 5th level: the student is able to use both conceptual and procedural knowledge to solve complex mathematical problems. He/she can recognize and use mathematical models to solve real-life problems, using different interpretations of the same mathematical object. He/she can find implicit data and information and is capable of a deep and correct mathematical reasoning.

In Table 1 (Patini-Liberatore, 2017) we present a model of converting the INVALSI results in the cognitive levels and difficulty of questions (*z=*z-points, *R=*results on the national assessment INVALSI, *l=*level of knowledge, *D=*difficulty of the question):

Table 1: Levels and results on the INVALSI.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *z* | <-2.00 | -2.00-1.50 | -1.50-1.00 | -1.00-.50 | -.50-0.0 | 0.00-.50 | .50-1.00 | 1.00-1.50 | 1.50-2.00 | >2.00 |
| *R* | <120 | 120-140 | 140-160 | 160-180 | 180-200 | 200-220 | 220-240 | 240-260 | 260-280 | >280 |
| *l* | Extremely bad | | Low | | Middle | | Good | | Excellent | |
| *D* | Extremely easy | Very easy | Easy | Middle-Easy | Middle | | Middle-difficult | Difficult | Very difficult | Extremely difficult |

# **Empirical research**

## *Aims of the research*

The literature showed that grades are correlated to students’ achievements on national assessments (Brennan et al., 2001; Willingham et al., 2002; Bowers, 2011; Argentin and Triventi 2015). Researchers have found that students’ performances in standardized tests are lower than their school grades (Nordin et al., 2019; Felda, 2018; Gershenson, 2018; Hurwitz & Lee, 2018). For what it concerns the situation in Italy, however, little or no research has focused on the question, whether are high school mathematics grades inflated. Such research question is nevertheless important, since it has been proved that students receiving inflated grades could get the feeling that they mastered a piece of the program that they actually did not (Chowdhury, 2018; Finefter-Rosenbluh & Levinson, 2015; Butcher et al., 2014).

Hence, our research questions are the following:

RQ1: are students’ mathematics grades correlated to their achievements on the standardized mathematics test?

RQ2: is grade inflation (for what it concerns mathematics) present also in Italian high schools?

## *Methodology*

The methodology used in this research is the descriptive and inferential statistical method together with the non-experimental method for causal analysis. We decided to use such method, because it would give us the opportunity to answer our research questions.

## *Statistical sample*

The statistical sample was provided by the Institute INVALSI, i.e. it was accessible on the webpage of their statistical office, previous the registration to the site (INVALSI, 2019).

The sample was composed by 35802 grade 10 students form different Italian high schools. There were 18479 (51.61%) female students.

At the end of the first semester, students normally get two grades for mathematics: an oral and a written one. Teachers calculate the written grade by considering written tests (i.e. “*compiti in classe*”), while for the oral grade, they include grades from oral interrogations, homework, and other tests, which are not “*compiti in classe*”. In order to analyze the school grades, we considered only the oral grades for mathematics, since the written mark is present only in some school typologies (e.g. high schools specializing in scientific education or some technical schools). We hereby excluded from our sample all those students who did not have an oral grade in mathematics. We did exclude 3427 students; our final sample had 32365 students, where 16898 (52.21%) were girls.

From the sample it was impossible to determine the average age of students, since for some students it was indicated, that they were born in “2000 or sooner” or “2004 and later”.

## *Procedure*

Firstly, to answer our first research question, i.e. whether mathematics school grades are correlated to students’ achievements on the INVALSI tests, we used the Pearson’s correlation coefficient, as proposed by Argentin and Triventi (2015).

Secondly, to answer our second research question, i.e. whether grade inflation is present also in Italian high schools, we converted students’ achievements on the INVALSI mathematics test into school grades, as proposed by Felda (2018). To do so, we considered the idea behind the Patini-Liberatore (2017) table (see Table 1), assigning the first positive grade (i.e. “6”) to all the achievements from 200 to 220 excluded. The conversion table is present in Table 2.

Table 2: Conversion of points in the INVALSI assessment into school grades.

|  |  |  |  |
| --- | --- | --- | --- |
| Points | Level | Qualitative description | School grade |
| <120 | Not even 1 | Extremely bad | 1 |
| 120-139 | 1 | Very bad | 2 |
| 140-159 | 3 |
| 160-179 | 2 | Bad | 4 |
| 180-199 | 5 |
| 200-219 | 3 | Sufficient | 6 |
| 220-239 | 7 |
| 240-259 | 4 | Good | 8 |
| 260-279 | 9 |
| >280 | 5 | Excellent | 10 |

We used the *t*-test for paired samples to compare students’ school grades and their hypothetical grades.

## *Data analysis*

All the data was analyzed with the statistical software *Jamovi*.

# **Results**

Analyzing the data, we found that for the school grades in mathematics at the end of the first semester, the mean was *M=*6.15 (*SD=*1.44; *min=*1; *max=*10), while for the performances at the INVALSI assessment, the mean was *M=*206 (*SD=*39.1; *min=*72.3; *max=*314). There are 10187 (31.48%) students, who got a grade lower than 6, i.e. a “bad/insufficient” grade (see also Table 3).

Moreover, we found a positive and statistically significative correlation between school grades and performances at the national assessment, suggesting that students with higher school grades get higher points at the national assessment (Pearson’s *r=*.428; *p*<.001). Hence, we answered our first research question: mathematics school grades and INVALSI achievements are positively and statistically significantly correlated.

To find evidence of grade inflation, we calculated the average scores at the INVALSI for each school grade, the standard deviation, the maximum and minimum, and the range (see Table 3).

Table 3: Descriptives for school grades.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| School grade | | *f(%)* | *M* | | *SD* | *min* | *max* | *Range* |
| 1 |  | 3 (0.0%) | 148 |  | 12.7 | 134 | 159 | 24.8 |
| 2 |  | 114 (0.4%) | 168 |  | 30.0 | 82.5 | 259 | 176 |
| 3 |  | 827 (2.6%) | 173 |  | 29.3 | 75.5 | 263 | 187 |
| 4 |  | 3191 (9.9%) | 182 |  | 32.3 | 74.0 | 303 | 229 |
| 5 |  | 6052 (18.7%) | 193 |  | 34.8 | 72.3 | 314 | 242 |
| 6 |  | 9580 (29.6%) | 203 |  | 35.3 | 75.5 | 314 | 238 |
| 7 |  | 6781 (21.0%) | 214 |  | 36.4 | 73.4 | 314 | 240 |
| 8 |  | 4046 (12.5%) | 228 |  | 36.8 | 78.9 | 314 | 235 |
| 9 |  | 1600 (4.9%) | 244 |  | 37.6 | 133 | 314 | 181 |
| 10 |  | 171 (0.5%) | 251 |  | 38.4 | 151 | 314 | 163 |

We found that with the increase of the school grade, also the mean performance at the INVALSI increases, which is shown by the correlation coefficient. However, we also found that the range (i.e. the difference between the maximum and the minimum) is relatively high for each grade category. For instance, some students with the final school grade 10 would get a minimum of 151 points, which is approximatively the mean of students with final grades 1 or 2, suggesting the presence of grade inflation.

To verify the presence of grade inflation, comparing school grades with national assessments, we converted the performances into school grades by considering Table 2. We found that the mean of the hypothetical grading is *M=*5.78 (*SD=*1.93; *min=*1; *max=*10). With the *t*-test for paired samples we found that there are statistically significant differences between students’ mathematics grades and hypothetical grades (*t*(32364)=-21.3; *p*<.001), indicating that hypothetical grades are lower that students’ school grades.

We then compared students’ grade with those we obtained converting the results of national assessment in school grades (see Felda, 2018), obtaining the situation depicted in Table 4. We found that teachers’ grades do not always reflect the students’ competences assessed by the national assessment INVALSI. Students who got an excellent grade by their teachers do not reflect the same excellent performance in the INVALSI’s assessment.

For students, who got an excellent grade at the end of the first semester, 129 (75.44%) had a worse performance in the national assessment. Among these, 21 (12.28%) students would get a bad grade (insufficient). Hypothetical grades that these students would get range from a minimum of 3 to a maximum of 10. Similarly, grades of students with a 9 in their report cards, would get hypothetical grades from 2 to 10. Among them, 1044 (65.25%) students would get a lower grade, 247 (15.44%) the same grade and 309 (19.31%) the excellent grade 10. 217 (13.56%) students with a 9 as their teachers’ grade, would get a bad grade. Students with the grade “insufficient” (5) would get hypothetical grades from 1 to 10, among them there are 2401 (39.67%) students, who would get a higher grade. Similarly, also those students with a bad grade (4), would get hypothetical grades from 1 to 10, among which 1569 (44.94%) students would get a higher grade.

Table 4: Comparison of real grades with hypothetical “corrected” grades.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Corrected grade | | | | | | | | | |
| Grade | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0  (0.0%) | 1  (33.3%) | 2  (66.7%) | 0  (0.0%) | 0  (0.0%) | 0  (0.0%) | 0  (0.0%) | 0  (0.0%) | 0  (0.0%) | 0  (0.0%) |
| 2 | 4 (3.5%) | 17  (14.9%) | 29  (25.4%) | 26  (22.8%) | 17  (14.9%) | 17  (14.9%) | 3  (2.6%) | 1  (0.9%) | 0  (0.0%) | 0  (0.0%) |
| 3 | 30  (3.6%) | 70  (8.5%) | 175  (21.2%) | 235  (28.4%) | 175  (21.2%) | 91  (11.0%) | 36  (4.5%) | 14  (1.7%) | 1  (0.1%) | 0  (0.0%) |
| 4 | 72  (0.0%) | 165  (5.2%) | 548  (17.2%) | 837  (26.2%) | 716  (22.4%) | 458  (14.4%) | 237  (7.4%) | 106  (3.3%) | 44  (1.4%) | 8  (0.3%) |
| 5 | 75  (1.2%) | 245  (4.1%) | 725  (12.0%) | 1215  (20.1%) | 1391  (23.0%) | 1113  (18.4%) | 736  (12.2%) | 346  (5.7%) | 150  (2.5%) | 56  (0.9%) |
| 6 | 69  (0.7%) | 198  (2.1%) | 692  (7.2%) | 1573  (16.4%) | 2132  (22.2%) | 1929  (20.1%) | 1564  (16.3%) | 861  (9.0%) | 374  (3.9%) | 188  (2.0%) |
| 7 | 25  (0.4%) | 65  (1.0%) | 298  (4.4%) | 796  (11.7%) | 1288  (19.0%) | 1434  (21.2%) | 1233  (18.2%) | 893  (13.2%) | 465  (6.9%) | 284  (4.2%) |
| 8 | 2  (0.1%) | 15  (0.4%) | 65  (1.6%) | 302  (7.5%) | 554  (13.7%) | 794  (19.6%) | 823  (20.3%) | 688  (17.0%) | 435  (10.8%) | 368  (9.1%) |
| 9 | 0  (0.0%) | 3  (0.2%) | 9  (0.6%) | 72  (4.5%) | 133  (8.3%) | 225  (14.1%) | 301  (18.8%) | 301  (18.8%) | 247  (15.4%) | 309  (19.3%) |
| 10 | 0  (0.0%) | 0  (0.0%) | 3  (1.8%) | 3  (1.8%) | 15  (8.8%) | 14  (8.2%) | 28  (16.4%) | 40  (23.4%) | 26  (15.2%) | 42  (24.6%) |
|  | | | | | | | | | | |

From Table 4 we might observe that among the 171 excellent students (i.e. those with the oral grade 10), only 24.6% (42 students) actually wrote the INVALSI test for an excellent grade, while the other more than three quarters of students would get a grade that is lower than the teachers’. Similar considerations can be made also for other grade categories. Considering hence the presented results, clear differences between school grades and grades on the INVALSI can be seen, suggesting that grade inflation in present also in Italian high schools.

# **Discussion**

Since the problem of grade inflation has not been recently studied for the situation of Italian high schools, we wanted to replicate the results found in the international literature, while adding an original part regarding the conversion of the INVALSI points into school grades.

One of the first aspects we found was that school grades and achievements in the national assessment INVALSI are positively and statistically significantly correlated (*r=.*428), which answered our first research question: students with higher school grades do have (statistically) a better performance at the INVALSI. Our result is coherent with the international literature (Brennan et al., 2001; Willingham et al., 2002; Bowers, 2011), however we did find a higher correlation than Argentin and Triventi (2015). Previous researches in Italy considered the whole population, while we considered a sample.

We furthermore found that the differences between the mean scores and the minimal ones are relatively big. We found that students, who got an excellent school grade, scored in the range from 151 to 314 points at the INVALSI (*M=*251; *SD=*38.4). The lowest achievement does differ from the mean for 2.6 standard deviations. Similarly, the lowest performance for students with a 9 at the end of the first semester (*M=*244; *SD=*37.3) was 133, which differs from the mean for 2.98 SDs.

The data we found suggest that teachers’ grades do differ from the performances at the national assessment of knowledge. Good students also score very low at the INVALSI.

These differences could derive from the problem of grade inflation. In order to show how this phenomenon is present in Italian high schools, we developed a grading system which was based on the INVALSI’s metric (INVALSI, 2018). Our aim was to show that good and excellent students would get hypothetical grades that are lower than those their teachers gave them in their report cards. We found that three quarters of students, who got an “excellent” (10), would get a lower grade. While considering the results of the INVALSI assessment as an objective metric of students’ knowledge, these students had inflated grades at the end of the first semester. Of course, this conclusion in based by comparing teachers’ grades with those we hypothetically proposed. There are several other factors which could explain these differences, such as test anxiety, cheating etc. Additional research is required in order to verify the impact grade inflation has in Italian high schools, however this initial study has shown a significant correlation between test performance and grades inflation.

Another important point that this study raises is related to teaching practice. In the Italian grading system, if a student gets a grade lower than 6 (i.e. a “negative” grade) at the end of the second semester, he/she would get a resit exam which is held normally in August or September, when the student should show that he/she did understand the studied topics. Hence, the “resit exam” is a way of giving the student an additional chance of passing the year, while getting some additional help from the teachers and colleagues.

In our sample, only grades at the end of the first semester were known. We found 10187 (31.48%) students with a grade lower than 6. From the proposed grading system, which considered the results in the national assessment, the total number of “bad” students would increase to 15082 (46.60%). The grading system we proposed is stricter than the standard teachers’ grading system: 15.12% more students would indeed have the “resit exam” if the grades at the end of the semester would also reflect the final performance in mathematics. We, therefore, might conclude, that many students would not have been provided with the additional help they would need. We found that 15347 (47.42%) students would get a grade lower than the one they got in their report cards.

On the other hand, our work showed that 10459 (32.32%) students would get a higher grade than the one given by the teachers, while 6559 (20.27%) students did get the same grade their teacher gave them.

From the obtained results we might conclude that the results in the national assessment INVALSI do differ from teachers’ grading system. Hypothetical grades that students would get, whether we did consider solely their performances in the INVALSI, suggest that grade inflation is a concerning issue in Italian secondary schooling. Whit grade inflation teachers could deprive students, who would need a clear feedback and information about their difficulties, of additional help (i.e. “resit exams”). Inflation of school grades could indeed harm students, schools and the national schooling and education system (Gershenson, 2018; Finefter-Rosenbluh & Levinson, 2015). Having a good understanding of students’ competences and knowledge is of paramount importance if we are to help those who are weaker and have some difficulties in mathematics. Getting higher grades could give students and parents a false sense of mastery and understanding, while excluding the students from additional help (Gershenson, 2018; Finefter-Rosenbluh & Levinson, 2015).

INVALSI assessment cannot however measure students’ total knowledge and performances, since (Quadro di Riferimento, 2018):

* it cannot assess students’ metacognitive and non-cognitive achievements, such as students’ attitude towards mathematics, or hard-working;
* cannot assess students’ usage and understanding of symbolic language, students’ proving of theorems, logical reasoning;
* do not consider conative aspects of the learning of mathematics.

Moreover, INVALSI tests are mainly composed of closed-type questions (Quadro di Riferimento, 2018), which have several limitations in assessing student’s knowledge. For example, guessing the correct answer is a concerning issue (see e.g. Andrich et al., 2016), since students have 25% of probability of guessing the correct answer. Also, the issue of cheating while taking the test INVALSI has been widely discussed (Longobardi et al., 2018). Hence, some high results in the INVALSI tests could be due to cheating or guessing. Hereby, students’ grading should consider both objective results (i.e. those given by national assessment of knowledge) and teachers’ opinions (Gershenson, 2018).

# **Conclusions**

With our empirical research we aimed to understand whether grade inflation is also present in Italian high schools, since the problem has been studied very little in this country.

We found a positive and statistically significant correlation between school grades and performances in the national assessment of mathematical knowledge INVALSI. We found, however, that relatively big differences between the average scores and the minimal scores are present among different grade categories. This indicated that some students, who received a certain grade form their teachers, had relatively very poor performances in the national assessment, which confirms results found in other international studies as seen in the international literature.

To study the phenomenon of grade inflation in Italian high schools, we developed a grading system based on the INVALSI metric. We found that the hypothetical grades differ from teachers’ grades, especially for excellent and good students. We found that more than 75% of excellent students would get lower grades, some of them even bad grades. Moreover, with our grading system we proved that the number of bad students would increase by 15%, which indicates that there are many Italian students who would need additional help (e.g. the “repairing exam”). With grade inflation, some teachers could deprive students of such help, giving them higher grades and false information about their real competences.

National assessments INVALSI cannot, however, give a complete view of students’ knowledge and non-cognitive skills, hence we propose a hybrid way of grading students, which takes into consideration both objective national assessment results and teacher grades.

# **References**

Ali, H., Ullah, S., ul Kashif, N., and Hussain, B. (2016). Analysis of Grade Inflation at Secondary School Level: Case Study of Board of Intermediate and Secondary Education, Multan. *Pakistan Journal of Social Sciences, 36*(2), 1011-1022.

Andrich, D., Marais, I., and Humphry, S. M. (2016). Controlling guessing bias in the dichotomous Rasch model applied to a large-scale, vertically scaled testing program. *Educational and psychological measurement, 76*(3), 412-435.

Argentin, G., & Triventi, M. (2015). The North-South divide in school grading standards: New evidence from national assessments of the Italian student population. *Italian Journal of Sociology of Education, 7*(2), 157-185.

Ayubbuzder, M., and Ali, A. (2013). Assessment of students’ learning achievements under semester system in Pakistan. *Journal of Basic and Applied Scientific Research, 3*(6), 79-86.

Babcock, P. (2010). Real costs of nominal grade inflation? New evidence from student course evaluations. *Economic inquiry, 48*(4), 983-996.

Bachan, R. (2017). Grade inflation in UK higher education. *Studies in Higher Education, 42*(8), 1580-1600.

Bansilal, S. (2017). The difficulty level of a national assessment of Grade 9 mathematics: The case of five schools. *South African Journal of Childhood Education, 7*(1), 1–8.

Bar, T., Kadiyali, V., & Zussman, A. (2009). Grade information and grade inflation: The Cornell experiment. *Journal of Economic Perspectives, 23*(3), 93-108.

Bauer, P., and Sheldon, G. (2008). *Ethnic discrimination in education: the Swiss Case.* WWZ Working Paper, No. 08/08, University of Basel, Center of Business and Economics (WWZ), Basel. http://hdl.handle.net/10419/123393. Accessed: 10 December 2020.

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159.

Brennan, R., Kim, J., Wenz-Gross, M., & Siperstein, G. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard educational review, 71(*2), 173-217.

Butcher, K. F., McEwan, P. J., and Weerapana, A. (2014). The effects of an anti-grade-inflation policy at Wellesley College. *Journal of Economic Perspectives, 28*(3), 189-204.

Caruth, D. L., and Caruth, G. D. (2013). Grade inflation: an issue for higher education?. *Turkish Online Journal of Distance Education, 14*(1), 102-110.

Chowdhury, F. (2018). Grade Inflation: Causes, Consequences and Cure. *Journal of Education and Learning, 7*(6), 86-92.

Costley, K. C. (2014). Stop Giving in to Higher Grades: Ten Suggestions on How to Fight Grade Inflation. https://files.eric.ed.gov/fulltext/ED546783.pdf. Accessed 18 December 2020.

Felda, D. (2018). Preverjanje matematičnega znanja. *Revija za elementarno izobraževanje, 11* (2), 175–188.

Feron, E., Schils, T., and Ter Weel, B. (2016). Does the teacher beat the test? The Value of the Teacher’s Assessment in Predicting Student Ability. *De Economist, 164*(4), 391-418.

Finefter-Rosenbluh, I., and Levinson, M. L. (2015). What is wrong with grade inflation (if anything)?. https://dash.harvard.edu/bitstream/handle/1/32228641/Whats\_Wrong\_With\_Grade\_Inflation.pdf?sequence=1. Accessed 26 December 2020.

Gershenson, S. (2018). Grade Inflation in High Schools (2005–2016). https://fordhaminstitute.org/sites/default/files/20180919-grade-inflation-high-schools-2005-2016\_0.pdf. Accessed 15 December 2020.

Hurwitz, M., and Lee, J. (2018). *Grade inflation and the role of standardized testing*. Measuring success: An examination of the use of standardized tests in college admissions. Baltimore, MD: Johns Hopkins University Press, p. 64-93.

INVALSI (2018). https://invalsi-areaprove.cineca.it/docs/2018/Livelli\_INVALSI\_g8.pdf. Accessed 10 December 2020.

INVALSI (2019). https://invalsi-serviziostatistico.cineca.it/. Accessed 8 December 2019.

Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied economics, 45*(9), 1185-1200.

Kartianom, K., Mardapi, D. (2017). The utilization of junior high school mathematics national examination data: A conceptual error diagnosis. *Research and Evaluation in Education, 3*(2), 163–173.

Keller, T. (2016). Sticky assessments–the impact of teachers’ grading standard on pupils’ school performance. *Educational Studies, 42*(5), 493-518.

Khan, E. A., and Hussain, B. (2019). Grade Inflation in Universities of Pakistan: A Case of Bahauddin Zakariya University, Multan. *Pakistan Journal of Social Sciences, 39*(3), 1037-1047.

Kunnath, J. P. (2017). Teacher Grading Decisions: Influences, Rationale, and Practices. *American Secondary Education, 45*(3), 68-88.

Lindahl, E. (2007). *Comparing teachers’ assessments and national test results: evidence from Sweden*. https://www.ifau.se/globalassets/pdf/se/2007/wp07-24.pdf. Accessed: 20 December 2020.

Longobardi, S., Falzetti, P., and Pagliuca, M. M. (2018). Quis custiodet ipsos custodes? How to detect and correct teacher cheating in Italian student data. *Statistical Methods and Applications, 27*(3), 515-543.

Nordin, M., Heckley, G., and Gerdtham, U. (2019). The impact of grade inflation on higher education enrolment and earnings. *Economics of Education Review, 73*, 101936.

Olutola, A. T., Olatoye, R. A., and Owolabi, H. O. (2019). Predictive Validity of Students’ Final Continuous Assessment Scores in Mathematics and Biology using their Performance in National Examination Council as a Criterion Measure. *FUDMA Journal of Educational Foundations, 1*(1), 147-158.

Özdemir, M., and Özkan, Y. (2017). Approaching to teachers’ grading: what is the correct order for implementing peer-and self-assessment in higher education context. *INTED2017 Proceedings*. IATED Academy, Valencia, Spain.

Patini-Liberatore (2017). http://www.patiniliberatore.edu.it/pvw/app/AQII0014/pvw\_img.php?sede\_codice=AQII0014&doc=2093906. Accessed: 25 January 2021.

Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher, 42*(5), 259-265.

Pramudita, K., Rosnawati, R., and Mam, S. (2019). Methods used by mathematics teachers in developing parallel multiple-choice test items in school. *Research and Evaluation in Education, 5*(1), 10-20.

Quadro di Riferimento (2018). https://invalsi-areaprove.cineca.it/docs/file/QdR\_MATEMATICA.pdf. Accessed: 12 December 2020.

Rapporto (2019). https://invalsi-areaprove.cineca.it/docs/2019/Rapporto\_prove\_INVALSI\_2019.pdf. Accessed: 8 December 2020.

Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, 22*, 1-38.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*(1), 1-37.

Wikström, C., & Wikström, M. (2005). Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. *Economics of education Review, 24*(3), 309-322.

Ziomek, R. L., & Svec, J. C. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin, 81*(587), 105-113.

**Abstract:** In this paper we analyze the problem of grade inflation in Italian high schools through an analysis of the results on the national assessment of knowledge, the INVALSI. Comparing the hypothetical grade students would receive on the national assessment with their school grades, we found signs of grade inflation. Almost 50% of students would get a lower grade if results from the INVALSI would have been considered. Among excellent students, more than 75% would get a grade lower than excellence. In the paper we argue that grade inflation could lead to a deprivation of help for those students, who would need it.

**Key words:** evaluation; grade inflation; high school; INVALSI; national assessment of knowledge.

**Povzetek:** V prispevku analiziramo problem inflacije ocen v italijanskih srednjih šolah preko analize učenčevih dosežkov na nacionalnem preizkusu znanja INVALSI. Primerjali smo hipotetične ocene, ki bi jih dijaki dobili na INVALSI-ju, z ocenami iz matematike, ki so jih dijaki dobili v spričevalu. Analiza kaže na prisotnost pojava inflacije ocen, saj bi skoraj 50 % učencev dobilo nižjo oceno od dejanske. Več kot 75% odličnih dijakov bi dobilo nižje ocene. V prispevku predstavimo težave, ki jih prinaša inflacija ocen, kot je na primer pomanjkanje dodatne pomoči tistim dijakom, ki bi jo potrebovali.

**Ključne besede:** ocenjevanje; inflacija ocen; srednja šola; INVALSI; nacionalno preverjanje znanja.