Double Standards in Educational Standards – Do Schools with a Disadvantaged Student Body Grade More Leniently?[a]

Oliver Himmler
Max Planck Institute, Bonn[b]

and

Robert Schwager
Georg-August-Universität Göttingen[c]

Abstract

A simple model of decentralized graduation standards is presented. It is shown that a school whose students are disadvantaged on the labor market applies less demanding standards because such students have lower ability or less incentives to graduate. The model's predictions are tested using Dutch school-level data. Since students in the Netherlands have to participate in both a central and a school specific examination, we can identify the grading policy of individual schools. We find that schools which harbor greater shares of disadvantaged students tend to set lower standards. This association is most pronounced in the track of secondary schooling preparing for university.

Keywords: education, grading, social status, schools, Netherlands
JEL: I21, J15

---

[a] We are grateful to the *Deutsche Forschungsgemeinschaft* for funding part of this research in the priority programme 1142 "Institutional Design of Federal Systems: Theory and Empirical Evidence". We thank Fabienne Plieschke for her assistance in assembling the data. We are indebted to Thiess Büttner, Dennis Epple, Volker Meier, and two anonymous referees for helpful comments. All remaining errors are our own.

[b] Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, phone: +49/228/91416-68, email: himmler@coll.mpg.de

[c] Department of Economics, Platz der Göttinger Sieben 3, D-37073 Göttingen, phone: +49/551/39-7244, email: rschwag@uni-goettingen.de

1 Introduction

Education policy is widely seen as a means to promote social policy goals. Good schooling is supposed to help the children of disfavored members of society to earn higher incomes than their parents and to gain social status. In most countries, however, students do not reap the benefits of formal education just by attending. Instead, they need to obtain the appropriate degree, preferably with good grades. Therefore, any social impact of education policy is filtered through the grading and examination system. Whether good schools will contribute much to social mobility depends on the way standards are chosen, and whether this choice depends on the social origin of students. In this paper, we therefore examine, both theoretically and empirically, the interaction of the social status of a school's students and the standards applied at examination.

In the first part of the paper we introduce status into a model of the choice of examination standard provided by Costrell (1994, 1997). In this model, each school sets its graduation standard so as to maximize the sum of the wages earned by its students. This decision is governed by the trade-off between the number of graduates, which decreases if the standard is more demanding, and the wage earned by each graduate, which increases in the standard. We extend Costrell's formulation by assuming that, in addition to the standard, also the social origin affects the wage earned by graduates. For a given standard, students from disadvantaged backgrounds obtain a lower wage than students from other social classes. We show that in this setup, schools with a disadvantaged student body set lower standards than other schools. Standards are inflated in this way because the wage discount experienced by graduates from unfavorable backgrounds depresses the return to learning effort for these students. They are thus less willing to satisfy any given standard than students from average social origins. To make up for the resulting loss in the number of graduates schools with disadvantaged students choose less demanding standards. A similar result obtains if the disfavored students have, on average, lower ability than those of other origins. Based on the above argument, however, our analysis shows that such an ability differential is not needed to explain lower standards in schools with a disfavored student population.

If the standard applied by a single school is not observable by employers, the graduates from several schools are pooled together in a common labor market, earning the same wage. We show that in such a scenario the equilibrium standard is decreasing in the size of the relevant labor market, that is, the number of schools whose graduates are pooled together. Our model thus confirms the well-known grading externality induced by locally determined but unobservable standards: Schools have an incentive to free-ride on high wages brought about by the other schools' tough grading. This mechanism has an implication for social policy, which is our focus here. It is plausible that the students from different social backgrounds are not equally mobile when applying for jobs. Specifically, it may be that disadvantaged students on average stay closer to their original residence. We show that, if this is true, the externality will be smaller in the case of disadvantaged schools, counteracting the tendency to set lower standards induced by unequal job prospects.

In the second part of the paper, we test the theoretical model using data from the Netherlands. This choice of subject is motivated by several features of the Dutch education system. Most importantly, students must pass central exams as well as school specific exams in order to receive a diploma. Thus, we are able to use the grades earned in the central examination as a benchmark against which to measure standards employed by individual schools in the school specific examination. In addition, secondary education in the Netherlands is organized in several tracks directed towards different further careers, from pre-university education to practical vocational training. This allows us to differentiate our analysis of grading standards according to different labor markets targeted by the different tracks.

The empirical analysis aims at explaining differences in standards chosen by different schools. In order to do so, we use the difference between the average grade of the school specific and the central examination as the dependent variable. The key explanatory variables are two proxies for the social status of a school's students, the percentage of cultural minority students and the percentage of students eligible for financial aid. It turns out that in most specifications these variables indeed show a significantly positive association with the grade difference. Thus, the empirical analysis is in line with the main prediction of the theoretical model: Schools with a higher percentage of disadvantaged students use a more lenient grading scheme than other schools. Leniency is largest for the school track which leads to university, smaller but still significant for the second highest academic level, and insignificant in the case of the more practically oriented tracks. Since it is plausible that graduates of higher academic tracks are more mobile, we take these results as an, albeit weak, evidence of the importance a school's market size has for the grading standard applied.

The paper contributes to the broad literature in education economics which analyzes the effects of the social and ethnic composition of schools (Epple and Romano, 1998; Nechyba, 1999; Epple, Newlon, and Romano, 2002; Hanushek, 2002; Entorf and Minoiu, 2005). More specifically, our work is related to several studies analyzing examination standards. Most of this research is concerned with the impact of different institutional arrangements for testing and examining students on student achievement. It has been well established by this strand of research that central standard setting in education paired with centrally devised and graded examinations leads to higher achievement in standardized tests. Theoretical foundations for this claim can be found in Costrell (1997) and Jürges, Richter, and Schneider (2005). Empirically, the performance enhancing effect of central standards has been established, among others, by Bishop et al. (2001), Betts (1998b), and

Jacob (2005) for the United States,[1] and by Bishop (1997, 1999), Wößmann (2003), and Jürges, Schneider, and Büchel (2005) for other countries.

Much rarer are studies which aim at explaining how and why standards are set. The basic theoretical approach followed in the present paper has been advanced by Betts (1998a) and Costrell (1994, 1997). More recently, Chan, Li, and Suen (2007) endogenize pooling across several types of schools in a signalling model of grading standards. None of these papers, however, addresses the issue of students' social origin.

This topic is treated by Burgess and Greaves (2009). Using data from British examination records, these authors compare student grades in central examinations with teacher assessments of student abilities. They find that students belonging to some ethnic minorities (such as black Caribbean) are more likely to be downgraded by the teacher relative to their performance in the central exam than students from the (white) majority. In a similar spirit, Prenzel at al. (2005), Kiss (2010), and Lüdemann and Schwerdt (2010) conclude that immigrant or lower class children in Germany receive lower grades, and are less likely to attend the branch of secondary schooling preparing for university than natives from higher classes, even after accounting for individual ability. Since theses studies analyze different countries and a different kind of data, it is difficult to discern the origin of these apparently diverging findings. Nevertheless, in the concluding section, we provide and discuss some possible explanations.

Wikström and Wikström (2005) analyze the determinants of grading standards in Sweden. Their approach is similar to ours since it also uses a central test as a benchmark against which local grading is measured. The Dutch central examination differs from this test, and is possibly more attractive as a benchmark, since it is compulsory, avoiding self-selection issues, and since its grades are measured on the same scale as the score of the local examination. Moreover, we directly address differences in the characteristics of

[1]See also the survey by Betts and Costrell (2001).

the schools' student populations, whereas Wikström and Wikström focus on competition among schools – a variable which differs across municipalities. They find that competition among public schools is associated with slight grade inflation. The paper also provides evidence that private schools competing with public schools engage in serious grade inflation.

The remainder of the paper is organized as follows. The following Section 2 contains the theoretical analysis. In Section 3 we give a brief overview of the institutional setup of the Dutch education system, describe the data, and present the estimation approach. Section 4 then contains the results of the empirical analysis. The concluding Section 5 discusses how our results compare to those found in Germany, and outlines possible future lines of research.

2 A Model of Standards and Social Status

In our model, schools set graduation standards which determine wages, and students choose how much learning effort to expend. Students have identical preferences over the wage $w \geq 0$ they will receive after leaving school and the learning effort $e \geq 0$ they expend at school. The effort is meant to reflect both time spent and the intensity of learning. The utility function is given by $u(w, e) = w - e^{\eta}$, where e^{η} describes the cost of learning effort, with a constant elasticity $\eta > 1$. Students differ in their ability to transform effort into examination results, as expressed by a student's learning productivity γ. The performance of a student at the examination is γe, and the standard set by the school is denoted by $s \geq 0$. A student with learning productivity γ who expends effort e graduates if and only if $\gamma e \geq s$.

Employers only observe whether a student graduates or not, whereas the actual examination performance γe, the learning productivity γ, and the effort e are private information of the student. By consequence, wages for graduates and non-graduates may differ, but wages cannot be conditioned on γ, e, or γe. In such a situation there is no reward to a student for exceeding the standard required for graduation. By consequence, a student with learning productivity γ will either expend just enough effort to satisfy the standard, $e = s/\gamma$, or she will dispense no effort at all, $e = 0$, and fail at the examination.

The wage received by non-graduates is normalized to zero. Denoting by $\tilde{w}$ the wage which a graduate from a given school may expect in the labor market, for a student of this school graduation is worthwhile if $\tilde{w} - (s/\gamma)^\eta \geq 0$. For any standard s and expected wage $\tilde{w}$, solving this condition for γ yields the graduation threshold $\gamma(s, \tilde{w}) = s\tilde{w}^{-1/\eta}$. All students whose learning productivity is at least as high as the graduation threshold, $\gamma \geq \gamma(s, \tilde{w})$, will graduate, and all those with $\gamma < \gamma(s, \tilde{w})$ will not.

Each school has an equal number of students, normalized to unity. There are two sets of schools $C = L, H$, where we denote also the numbers of the schools in both sets by L and H. The set L (H) contains schools with a student body originating from a disadvantaged (favored) social background. As a convenient, if over-simplifying, label we call the former the "lower-class schools" and the latter the "higher-class schools". For example, such social segregation in schools may be the result of Tiebout sorting in the local property market combined with substantial costs of commuting to schools located far away from the student's residence. The sets L and H are interpreted as containing all schools with a given social background which supply graduates to the same regional labor market. As an interesting and plausible case, we specifically consider the possibility that lower class workers are less mobile than higher class workers. Then the relevant labor market is smaller for lower class schools than for higher class schools, i.e., $L < H$.

At all schools $i \in C, C = L, H$, the ability parameter γ is distributed according to a uniform distribution over the interval $[0, \overline{\gamma}_C]$. Using this notation and the graduation threshold, one finds the number of graduates from school $i \in C$ when it sets a standard

s_i and when its graduates expect a wage $\tilde{w}_i$. Thus, for $0 \leq \gamma(s_i, \tilde{w}_i) \leq \overline{\gamma}_C$, the graduation rate from school $i \in C$ is given by $1 - \gamma(s_i, \tilde{w}_i)/\overline{\gamma}_C$. In this formulation, the upper bound $\overline{\gamma}_C$ provides an aggregate measure of the ability of the students attending a school in social class $C = L, H$. We allow these measures to differ between both sets of schools. Specifically, $\overline{\gamma}_L < \overline{\gamma}_H$ describes the case where students of disadvantaged backgrounds enter school with lower ability than students from favored origins.

Conditional on the standard s_i required by a school $i \in C, C = L, H$, employers' willingness to pay for a graduate from school i is $\lambda_C s_i$. This formulation expresses the idea that the examination performance s_i determines productivity at work, which for simplicity is measured in the same units. Moreover, social origin affects the wages according to the parameters λ_C, where we assume $0 < \lambda_L \leq \lambda_H = 1$. The parameter $\lambda_L \leq 1$ may reflect properties of disadvantaged students which are relevant for their productivity at the workplace but not tested in the examination, for example good manners, rhetorical abilities, stable families, belonging to social networks, or all sorts of "soft skills". As an alternative interpretation, λ_L might be identified with outright discrimination against disadvantaged workers in the sense that they are being paid less than workers from favorable origins in spite of identical productivity.[2]

Employers do not observe the standard s_i required by an individual school but they observe the social origin of the school's students. Such an informational scenario will occur, for example, if the residences of disadvantaged students are clustered in space so that the location of a school contains information about the social background of the school's students. At the same time, it may be too costly for firms to monitor the grading standards of individual schools. Consequently, wages may differ between higher-class

[2]Discrimination is difficult to rationalize when firms maximize profits. Since we do not explicitly describe firms' hiring choices, our model does not rule it out, however. Kee (1995) and Zorlu (2002) present empirical evidence that wage discrimination against ethnic minorities is present in the Netherlands. Similarly, van Ours and Veenman (2002) find that second generation immigrants in the Netherlands are less likely to be employed, conditional on education.

and lower-class schools but not according to the graduation standards of the individual schools. Following this argument, we assume that the wage paid to graduates from any school $i \in C, C = L, H$, is given by

$$w_C = \sum_{i \in C} \lambda_C s_i \frac{1 - \gamma(s_i, \tilde{w}_i)/\overline{\gamma}_C}{\sum_{j \in C} \left[1 - \gamma(s_j, \tilde{w}_j)/\overline{\gamma}_C\right]} \,. \tag{1}$$

That is, the wage is the weighted average of the wages which would, under full information, be paid to the graduates from the schools in the relevant labor market, where the weights are given by the shares of the individual schools in the total number of graduates.

In an equilibrium the wage is correctly anticipated by students when they choose their effort levels. Thus, for any vector of standards $(s_i)_{i \in C}$, an equilibrium wage is a fixed point of (1) satisfying $w_C = \tilde{w}_i$ for all $i \in C$.[3] It is determined implicitly by the condition

$$\sum_{i \in C} \Big(w_C - \lambda_C s_i\Big) \left[1 - \frac{\gamma(s_i, w_C)}{\overline{\gamma}_C}\right] = 0 \,, \tag{2}$$

which is derived by inserting $w_C = \tilde{w}_i$ for all $i \in C$ in (1). In the following, we focus on symmetric equilibria where all schools $i, j \in C$ of one class choose identical standards $s_i = s_j = s_C$, implying an identical graduation threshold $\gamma_C = \gamma(s_C, w_C)$. Then the equilibrium wage is uniquely determined and equal to $w_C = \lambda_C s_C$ for all schools in C. Moreover, starting from symmetric standards, a marginal change in standard by school i affects the equilibrium wage in class $C = L, H$ according to $\mathrm{d}w_C/\mathrm{d}s_i = \lambda_C/C$.[4] Intuitively, since there are C schools pooled in the labor market, an individual school's weight in determining the wage is only $1/C$.

[3] Since for all expected wages $(\tilde{w}_i)_{i \in C}$, the right hand side of (1) is just a weighted average of the values $\lambda_C s_i$ for all schools, for all vectors of standards such a fixed point exists in the interval $\left[\lambda_C \min_{i \in C}\{s_i\}, \lambda_C \max_{i \in C}\{s_i\}\right]$.

[4] To derive this, differentiate (2) implicitly and use the implications of symmetry mentioned above.

We assume that each school maximizes the sum of the wages earned by its students. Schools thus care for their students, without however taking effort costs into account.[5] This objective function can be motivated by observing that good career prospects of students contribute to a school's reputation, and thus may be useful for attracting students or grants. When deciding about the standards they require for graduation, schools anticipate the optimal choices by students and the equilibrium wage. If school $i \in C$ sets standard s_i, it thus expects that the wage for graduates from class C will be w_C according to (2), taking the standards $(s_j)_{j \in C, j \neq i}$ chosen by all other schools in the market as given. Since non-graduates earn a wage of zero, school i's maximization problem can thus be stated as

$$\max_{s_i \geq 0} \quad W_i(s_i) = \left[1 - \frac{\gamma(s_i, w_C)}{\overline{\gamma}_C}\right] w_C \, .$$

The necessary condition for an interior solution is

$$\frac{\partial W_i}{\partial s_i} = -\frac{w_C}{\overline{\gamma}_C}\left[\frac{\partial \gamma(s_i, w_C)}{\partial s} + \frac{\partial \gamma(s_i, w_C)}{\partial \tilde{w}}\frac{\mathrm{d}w_C}{\mathrm{d}s_i}\right] + \left[1 - \frac{\gamma(s_i, w_C)}{\overline{\gamma}_C}\right]\frac{\mathrm{d}w_C}{\mathrm{d}s_i} = 0 \, . \qquad (3)$$

Equation (3) shows the trade-off faced by a school. On the one hand, as expressed by the first term in square brackets in (3), a more demanding standard decreases welfare by reducing the number of graduates. On the other hand, a higher standard raises the wage for graduates. This enhances welfare both directly, as measured by the last term in (3), and indirectly by increasing the number of graduates. This effect, which is formalized by the second term in the first square brackets in (3), counteracts the decline in the graduation rate triggered by the higher standard.

In a symmetric equilibrium where $s_i = s_C, \gamma_i = \gamma_C, w_C = \lambda_C s_C$ and $\mathrm{d}w_C/\mathrm{d}s_i = \lambda_C/C$

[5] This omission reflects current debates in education policy which do not seem to be very concerned about students enjoying insufficient leisure.

for all $i \in C$, condition (3) can be solved explicitly[6] for the graduation threshold and the standard:

$$\gamma_C^* = \frac{\overline{\gamma}_C}{C+1-1/\eta} \qquad \text{and} \qquad s_C^* = \lambda_C^{1/(\eta-1)} \left(\frac{\overline{\gamma}_C}{C+1-1/\eta} \right)^{\eta/(\eta-1)} \tag{4}$$

In (4), the limiting case $C = 1$ represents a market consisting of only one school, which is equivalent to a scenario where the employers have full information about the standards applied by each individual school. Inspecting the second equation in (4), it is obvious from $\eta > 1$ that $\mathrm{d}s_C^*/\mathrm{d}\lambda_C > 0$ and $\mathrm{d}s_C^*/\mathrm{d}\overline{\gamma}_C > 0$ for both $C = L, H$. Specifically, for the grading policy of lower class schools this implies:

Proposition 1 *The larger the wage discount for graduates from disadvantaged social backgrounds, and the lower the learning productivity of such students, the lower is the standard chosen by a school with students from lower social classes.*

This result shows that a school which cares about the incomes of their students will grade more leniently if its students are socially disadvantaged. As one may expect, such behavior may simply be the consequence of lower abilities on the part of students from lower social classes. Proposition 1, however, shows that more lenient grading may just as well be the rational reaction of a school to the unfavorable job prospects of its graduates.

As a second result, we see from (4) that $\mathrm{d}s_C^*/\mathrm{d}C < 0$. Hence:

Proposition 2 *A smaller market size C raises the standard s_C^*.*

This proposition illustrates the well-known grading externality among schools sharing a common labor market. If a lower class school $i \in L$ marginally lowers its standard the willingness to pay for a graduate from this school decreases by λ_L. Since school i has only weight $1/L$ in the group of lower class schools this translates only into a wage decrease of

[6] Details, and a proof of the second order condition, are available from the authors upon request.

λ_L/L. Schools therefore have an incentive to free ride on the high wages brought about by the tough standards of other schools, by grading leniently themselves. The result is a general devaluation of standards which is more pronounced the larger the market is.

3 Data and Estimation Approach

We now turn to the empirical analysis of the interaction of standards and the social composition of schools as summarized in Propositions 1 and 2. As an introduction, we give a brief account of the education system in the Netherlands.

3.1 The Dutch education system

Dutch compulsory education encompasses twelve school years. Primary education starts at age five and lasts eight years. In secondary education, parents may choose among three tracks (*opleidingen*):

(i) Pre-vocational secondary education lasts four school years, and comes in two flavors: *voorbereidend beroepsonderwijs, VBO* and *middelbaar algemeen voortgezet onderwijs, MAVO*. Because the latter is more theoretical than the former, we will label *VBO* lower secondary and *MAVO* middle secondary education. Most students move on to vocational training after graduation, but especially for *MAVO* graduates it is also not uncommon to proceed to higher secondary education.[7]

(ii) Higher secondary education (*Hoger Algemeen Voortgezet Onderwijs, HAVO*) amounts to five years of schooling and is aimed at preparing students for entry into higher

[7] In 1999 the *MAVO* track was renamed to *VMBO-GT* and *VBO* was renamed to *VMBO-BK*. The 2003 cohort thus includes the first graduates from these rebranded tracks, which are equally referred to as middle (resp. lower) secondary education.

professional education, which typically leads to a bachelor's degree. Graduates may also enroll in the fifth year of pre-university education or opt for vocational training.

(iii) Pre-university education (*Voorbereidend Wetenschappelijk Onderwijs, VWO*) is the highest form of secondary education in the Netherlands and encompasses six years of schooling. Graduates typically take up a university education.

Many Dutch schools offer more than one of these tracks, and often one school will provide access to all three. Typically, the different tracks will not be offered in the same geographical location – in that sense a dutch 'school' often consists of multiple 'sub-schools' (*vestigingen*). From now on, whenever we speak of a school, what we have in mind is a 'sub-school'. When we refer to the broader school definition, we will use the term 'institution'.

In all of the tracks students end their scholastic careers with central examinations, and so at first glance Dutch secondary education appears to be characterized by central standards. However, central exams (*centraal examen*) account for only half the final grade. The other half is determined via decentralized testing (*schoolexamen*), leaving grading and standard setting to a large extent at the individual school's discretion.[8]

The central exams are arranged and graded by the testing agency *CITO* (*Central Institute for Test Development*).[9] All students of the same track are faced with identical questions and grading is done by the agency. Decentralized testing takes place in all subjects, whereas a few subjects are exempt from centralized testing (physical education and the arts). The analysis in this paper is limited to subjects where both types of testing

[8]An early study by Dronkers (1999) is also concerned with the school and central exams and discusses possible causes for discrepancies between the two grades. The author discards the possibility that higher decentral grades are due to an easier curriculum for these exams. The differences in grading are believed to be rooted in regional, subject-specific as well as teacher-specific grading traditions. Other explanations given include relative grading, and an idiosyncratic "grading culture" of the individual schools.

[9]For further information, go to: http://www.cito.nl/com_assess_ex/nat_final_ex/eind_fr.html

are employed and school exam grades can thus be compared to central exam grades. An official body appointed by the Ministry of Education establishes the norms for the central exams. The school exams on the other hand are conducted and –more importantly– devised and graded by the local schools. There are, however, guidelines set by the department of education concerning the subject matter covered in school exams, to which schools must abide. To this end, the local schools set up "exam rules" (*examenreglement*), which establish the curriculum and required reading for the local exams. These rules need to be accredited by the central authorities and are accessible to the respective school's students. Nonetheless, it is obvious that in essence it is the individual school which sets the standard, at least within a certain range.

3.2 Estimation approach

Our empirical approach uses the co-existence of central and school-specific grades in order to detect differences in local standards. Let G_i^c denote the average central exam grade and G_i^s the average school exam grade in school i. Under coinciding central and school specific standards, we would expect $G_i^s = G_i^c$. An upward deviation of G_i^s from G_i^c then constitutes a local standard that falls short of the central standard and vice versa. Our operationalization for the standard s_i applied by school i is therefore the difference $\Delta G_i = G_i^s - G_i^c$ between the average grades obtained at this school in the school specific and in the central examination.

In order to explain the grade difference ΔG_i, we use the estimation equation:

$$\Delta G_i = \alpha + \mathbf{y_i} \cdot \beta + \mathbf{x_i} \cdot \rho + \epsilon_i, \tag{5}$$

where i denotes the individual school, $\mathbf{y_i}$ contains variables describing the student body social composition, $\mathbf{x_i}$ is a vector of control variables, and ϵ_i is the error term. We will focus on two variables which capture the school-level social composition $\mathbf{y_i}$:

(i) The percentage of cultural minority students.

(ii) The percentage of students receiving study cost allowance *(Tegemoetkoming studiekosten)*, eligibility for which implies that parents have a low disposable income.

In accordance with Proposition 1, we expect to find decreasing local standards with increasing school-level percentages of (i) and (ii). That is, if double standards are employed, we expect β to be positive.

According to Proposition 2, an increase in market size leads schools to set lower educational standards if firms cannot distinguish the standards set by individual schools in a market. Since firms might derive standards from school level information on average grades, which is available in the Netherlands, we cannot provide a direct test of this hypothesis. However, it is not obvious that firms actually retrieve and use such information, which may be costly to do. Therefore, studying the different tracks of secondary education might still shed some light on the idea. This is because it is reasonable to think of job market size to be increasing in the level of education. That is, the relevant labor market is smallest in geographical terms for students who have earned a diploma in lower and middle secondary education and largest for those who hold a pre-university diploma. Hence, we estimate (5) separately for all school tracks and suspect that the cutting of standards as measured by β to be largest in pre-university education and smallest in lower secondary education.

On a formal level, the continuous variable ΔG_i departs from the binary pass-fail standard s_i which is featured in the theoretical model. Since it is likely that a school which grades leniently also awards degrees more easily, it is, however, plausible that the factors determining graduation standards affect average grades in a similar way. Moreover, it is plausible that many employers and universities require a certain minimum grade from applicants whom they are willing to consider seriously. In such a case, this is the standard a student must meet, and the grading scale effectively determines a binary standard.

The difference ΔG_i cannot in itself be interpreted normatively. It does not say whether the school specific or the central standard is correct in the sense of measuring the "true" skill level of students. A positive ΔG_i might be a correction for an overambitious central standard rather than grade inflation by school i.[10] In this paper we will not, however, question the appropriateness of the central grading scheme and therefore accept it as the benchmark against which double standards are to be detected. This is justified by the main focus of our investigation. We are not primarily interested in grade inflation itself. It may well be that on a local level teachers tend to award higher or lower grades in general, say because school exams tend to be standardized in a different manner than central exams. We would then expect ΔG_i to be different from zero but constant across schools. Our focus, instead, is on an association of standards with social status. If standards are socially differentiated, ΔG_i will be systematically related to the social composition of the schools' student body, irrespective of the average deviation between school specific and central grades.

3.3 Data sources

We use data from four different sources. School-level data on student performance and social affiliation as well as school characteristics are taken from the Quality Cards for Secondary Education (*Kwaliteitskaart Voortgezet Onderwijs*). This dataset is available from the Netherlands Inspectorate of Education; it covers all Dutch institutions in secondary education and provides information on number of students, administrative form of the institution (private, denominational, public), the school tracks that can be attended,

[10] It can be ruled out, however, that school grading is conditional on central grades, as the *centraal examen* is the last exam of the entire school career.

average class sizes, subject-level average grades attained in school and central exams, the recommended type of secondary school based on student performance in primary education, the percentage of ethnic minority students, and the percentage of students with a study cost allowance.

The dependent variable ΔG_i is constructed from the performance data of the 2003 and 2004 Quality Cards. As the original file contains interdisciplinary average grades only as a mean of school and central grades, we calculated the average school specific (G_i^s) and central (G_i^c) grades covering all subjects by weighting the average school and central grades in each subject with the number of students that had actually taken part in the exams in that particular subject.

The percentage of cultural minority students is defined as the share of students in a given school branch who have a certain non-Dutch background.[11] Along with the percentage of students receiving study cost allowance we use this variable as a proxy for low social status, as neither the Inspectorate nor the individual schools collect detailed data on parents' socioeconomic status. At the end of primary school each student is given a non-binding advice by her teachers as to which school branch is deemed appropriate in secondary education. This advice can be interpreted as a proxy for incoming student ability. "Above advice" ("below advice") denotes a student attending a more (less) demanding branch than the one recommended. In pre-university education there are no "below advice" students as this is the highest possible track. In lower secondary education a student who is "above advice" has received a special education track recommendation. We also include controls for the percentage of students in ability-tracked classes in the second year

[11]In order to be officially considered part of a cultural minority in the Netherlands, students have to satisfy one of the following criteria: Both parents were born in (or have the nationality of) one of the republics of former Yugoslavia, Greece, Italy, Cape Verde, Morocco, Portugal, Spain, Tunisia or Turkey, Moluccan background, Surinamese, Antillean or Aruban background, Roma background, caravan dwellers, other non-European background and not having completed full primary education in the Netherlands, Eastern European background and not having completed two years of Dutch schooling.

of secondary education and the average class size.[12]

In addition to school level data, we use postcode 'status scores' as a proxy for the students' social background. The status scores are supplied by The Social and Cultural Planning Office of the Netherlands (*SCP*), a Dutch government agency. This index takes into account local characteristics such as mean education, mean income and average rents, among others. Postcode areas that have a low social status are denoted with values greater than zero, areas of higher status receive negative values. We match these scores with the schools' 4-digit postcodes taken from the Quality Cards. More postcode level controls come from the *Kerncijfers postcodegebieden 2003* as well as the *Kerncijfers wijken en Buurten 2001-2005*, published by the Dutch Office of Statistics (*CBS*). The postcode percentage of school-aged children is calculated from the dataset *Bevolking per 4-cijferige postcode 2004*, published by the *CBS*.

Since school-level financial endowment as well as characteristics of the teaching staff might influence social composition and standard setting, data from the series *Onderwijs in Cijfers (OIC, literally translated: 'Education Figures')* is used in the estimation, too. *Onderwijs in Cijfers* is published annually by the Dutch Ministry of Education and is intended to provide school managers with information on the above mentioned matters for all Dutch secondary schools.

Table 1 displays descriptive statistics for all right hand side variables.

[12] Ability-tracked in this context means that in the first two years of secondary education students attend classes with students from their chosen branch only, whereas non-tracked students attend classes together with students from other branches. After the second year of secondary education there are no mixed classes.

Table 1: Descriptive statistics.

	pre-university (N=635)		higher secondary (N=597)		middle secondary (N=802)		lower secondary (N=454)	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Above advice %	22.03	16.86	12.17	11.88	6.27	12.30	12.95	15.98
Below advice %	-	-	4.98	5.09	14.14	10.89	15.96	14.07
Minority students %	3.74	6.91	5.72	9.61	7.09	11.69	13.43	16.53
Study cost%	26.77	9.95	36.87	13.47	39.81	18.74	65.38	22.01
Tracked %	65	35.80	61.39	36.00	61.71	36.14	59.81	36.10
Class size	25.81	2.34	25.67	2.41	23.90	3.52	20.61	3.36
public school	.31	.46	.29	.45	.29	.45	.31	.46
No. Students (in 1000s)	1.17	.44	1.21	.42	.95	.54	.89	.57
Short term debt (balance share)	.30	.10	.30	.10	.30	.10	.30	.11
Long term debt (balance share)	.04	.07	.03	.06	.03	.06	.03	.06
Staff growth	.03	.05	.03	.05	.03	.05	.03	.05
No. students growth	.02	.04	.02	.04	.01	.04	.01	.05
Part time staff %	.38	.09	.36	.08	.34	.07	.33	.06
Status postcode	.01	1.07	.02	1.04	.13	.96	.32	.89
Avg. income postcode (in 1000s)	13.97	2.83	13.81	2.64	13.25	2.18	12.72	1.51
School aged postcode %	16.96	3.48	17.27	3.29	17.54	3.15	17.47	3.22
Population postcode (in 1000s)	8.43	4.23	8.63	4.58	8.46	4.38	8.49	4.36

4 Empirical Results

4.1 Determinants of standards in Dutch schools

Descriptive statistics for central exam grades, school exam grades, and the difference between the two, ΔG_i, are shown in Table 2 for the pooled classes of 2002 and 2003. Grades run from 10 to 100, and on the central exams students in all branches score grades ranging from the lower 50s to the mid 70s on average.[13] Grades awarded in school exams are on average higher than those awarded in central exams in all branches but lower secondary. Thus, the local school standards in these branches seem to be lower than the centrally devised standard. The difference is highest for pre-university education and smallest in lower secondary schools. The minimum values drop only slightly below the central exam in pre-university education at most, whereas in lower secondary education the grades awarded in school exams are in some cases much lower than the central exam grades – the

[13]Dutch grades usually range from 1 to 10, however in the data supplied by the Dutch Inspectorate of Education the grades are multiplied by 10, and we stick with this scale.

middle tracks lie in between. Altogether, it seems that the schools in the higher tracks of secondary education reduce the standards more than the lower track schools. This is consistent with the idea of geographically larger job markets being associated with lenient grading, presented in Proposition 2.

Table 2: SUMMARY STATISTICS: GRADES 2002/2003.

Variable	N	Mean	Std. Dev.	Min	Max
central exam grade					
pre-university (vwo)	635	64.20	2.75	53.01	71.98
higher secondary (havo)	597	62.78	2.15	54.24	69.06
middle secondary (mavo)	802	64.11	2.35	53.61	73.06
lower secondary (vbo)	454	64.76	3.68	55.59	78
school exam grade					
pre-university (vwo)	635	68.82	1.49	61.42	73.30
higher secondary (havo)	597	65.29	1.10	61.60	68.89
middle secondary (mavo)	802	65.53	1.76	59.79	73.26
lower secondary (vbo)	454	64.61	2.26	55.56	78
grade difference ΔG_i					
pre-university (vwo)	635	4.62	2.50	−2.09	16.44
higher secondary (havo)	597	2.51	2.24	−4.98	12.01
middle secondary (mavo)	802	1.41	2.48	−5.00	11.38
lower secondary (vbo)	454	−.15	3.32	−9.01	14.5

Average central and school exam grades in the four tracks of Dutch secondary education (at school-level). Dutch grading is from 1 to 10, yet in the Quality Cards data grades are multiplied by a factor 10, so that grades in our data can range from 10 to 100. ΔG_i is school exam grade minus central exam grade.

As the divergence between central and school grades is largest in pre-university education, in Table 3 we first investigate the correlates of local standards in the highest track of secondary education. The dependent variable in these OLS regressions is the track-specific difference ΔG_i, and specification (1) includes typical school-level variables only, (2) adds *Onderwijs in Cijfers* variables and (3) includes postcode-level data. In accordance with the theoretical predictions, specifications (1) to (3) show that a higher percentage of cultural minority students is associated with lower standards in local exams. The share of students eligible for study cost allowance does not seem to be linked to lenient grading. One reason is that a large percentage of minority students also qualify for study cost allowance, resulting in quite some overlap in the two variables. This becomes obvious when we interchangeably employ only one of these two proxies for social status. Omitting the

Table 3: Baseline estimates for pre-university education.

	(1)		(2)		(3)		(4)		(5)	
Above advice %	.0307***	(0.008)	.0315***	(0.007)	.0316***	(0.008)	.0427***	(0.008)	.0315***	(0.008)
Minority students %	.0895***	(0.029)	.1141***	(0.027)	.1117***	(0.028)			.1171***	(0.026)
Study cost %	.013	(0.014)	.0011	(0.013)	.0075	(0.013)	.0457***	(0.016)		
Tracked %	−.0094***	(0.003)	−.0098***	(0.003)	−.0107***	(0.003)	−.011***	(0.003)	−.0105***	(0.003)
Avg. class size	−.1056***	(0.040)	−.0844**	(0.041)	−.0854**	(0.042)	−.1208**	(0.048)	−.0893**	(0.043)
Public school	.3905	(0.254)	.2855	(0.263)	.362	(0.270)	.407	(0.290)	.3631	(0.270)
No. of students	.2161	(0.224)	.2616	(0.247)	.2895	(0.250)	−.0056	(0.273)	.2969	(0.251)
Short term debt			2.515***	(0.849)	2.165**	(0.847)	2.423***	(0.851)	2.12**	(0.846)
Long term debt			2.88**	(1.312)	2.201	(1.425)	1.592	(1.368)	2.223	(1.427)
Staff growth			.3588	(1.667)	.2222	(1.651)	.0833	(1.714)	.2512	(1.650)
No. students growth			−1.925	(2.187)	−1.913	(2.207)	−1.359	(2.200)	−1.867	(2.206)
Part time staff %			−.0336	(1.263)	−.3065	(1.368)	−.4076	(1.404)	−.3551	(1.362)
Status postcode					.3047	(0.187)	.3984**	(0.201)	.312*	(0.187)
Avg. income postcode					.1353*	(0.075)	.1765**	(0.080)	.1331*	(0.074)
Share school aged postcode					.0663*	(0.037)	.0565	(0.042)	.0654*	(0.037)
Population postcode					−.068***	(0.023)	−.0658***	(0.023)	−.0669***	(0.023)
Year 2003	−.1696	(0.108)	−.1308	(0.120)	−.162	(0.123)	−.3119***	(0.117)	−.1398	(0.118)
constant	6.306	(1.208)	5.106	(1.388)	2.791	(2.080)	2.829	(2.373)	3.108	(1.980)
N	759		649		635		635		635	
adj. R^2	0.208		0.237		0.257		0.203		0.258	

The dependent variable is the difference between the school grade and the central exam grade. All estimations are for the pre-university track. Standard errors in parentheses allow for clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

percentage of minority students from the estimation in specification (4) results in a considerably larger and significant coefficient on the share of study cost recipients. Dropping the share of study cost allowance recipients in specification (5) yields a similar effect on the minority share coefficient. This supports the idea that these variables overlap, yet in sum there seems to be a link between social status and grading standards that is not due to pure economic status as measured by the number of study cost allowance recipients.

Table 4: Baseline estimates. Higher, middle, lower secondary branches.

	higher secondary		**middle secondary**		**lower secondary**	
Above advice %	0.045***	(0.010)	0.045***	(0.010)	0.010	(0.011)
Below advice %	−0.003	(0.170)	−0.034***	(0.008)	−0.020*	(0.011)
Minority students %	0.058***	(0.018)	0.043***	(0.007)	0.012	(0.009)
Study cost %	0.007	(0.010)	0.002	(0.005)	0.008	(0.007)
Tracked	−0.004	(0.003)	−0.005**	(0.002)	−0.004	(0.004)
Avg. class size	0.032	(0.040)	−0.119***	(0.028)	−0.075	(0.053)
Public school	0.374	(0.243)	−0.132	(0.202)	−0.435	(0.291)
No. of students	0.286	(0.235)	0.030	(0.176)	0.234	(0.318)
Short term debt	0.484	(0.900)	0.262	(0.751)	−2.503**	(1.178)
Long term debt	1.303	(1.451)	−2.133	(1.506)	−2.805	(2.521)
Staff growth	−1.729	(1.632)	1.366	(1.366)	−1.796	(2.633)
No. students growth	−0.776	(2.138)	−0.047	(1.883)	−3.654	(2.984)
Part time staff %	0.635	(1.412)	−0.887	(1.101)	2.386	(2.423)
Status postcode	0.573***	(0.196)	0.292*	(0.167)	0.057	(0.309)
Avg. income postcode	0.276***	(0.076)	0.236***	(0.070)	0.251	(0.202)
Share school aged postcode	0.078**	(0.035)	0.010	(0.036)	0.022	(0.074)
Population postcode	−0.023	(0.021)	0.010	(0.020)	0.050	(0.032)
Year 2003	0.130	(0.116)	−1.010***	(0.122)	3.527***	(0.229)
constant	−5.004	(2.053)	1.740	(1.576)	−4.505	(4.197)
N	597		802		454	
adj. R^2	0.239		0.296		0.342	

Estimates are for the lower three secondary school tracks. Columns show estimates for the school branch sample indicated in the header. The dependent variable is the difference between the school grade and the central exam grade. Standard errors in parentheses allow for clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 extends the analysis to the other three school tracks, where column (1) describes grading in higher secondary education, column (2) shows results for middle secondary education, and column (3) for lower secondary education. The main result holds for higher secondary as well as middle secondary education: a lower status student body, as measured by the share of cultural minority students, is linked to depreciating standards. The coefficient for study cost allowance recipients is again not statistically significant. It becomes so upon omission of the cultural minority variable for lower and middle secondary education (not reported). Interestingly, the coefficients on the share of cultural minority students in higher and middle secondary education are only about half as large as in pre-university education. Moreover, the grade gap in lower secondary education does not at all seem to be related to social composition of the student body. Together, we take these results as tentative evidence in favor of the hypothesis that schools supplying graduates to smaller job markets tend to inflate disadvantaged students' grades by less.

Taken as a whole, the empirical results for the pre-university, higher and middle secondary tracks clearly reject the hypothesis that social composition is not related to the magnitude of standard cutting. The direction of the relationship does nothing to suggest that students in low status schools might be discriminated when it comes to local grading policies. If anything, the opposite appears to be true. Moreover, the size of the coefficients for the different tracks is consistent with the hypothesis that standard cutting increases in the relevant labor market size.

4.2 Disadvantaged students – Social status or ability?

The systematic differences in standards may be the consequence of diverging grading schemes or of a different choice of examination topics. In the first case, schools may have school examinations of equal difficulty, but those with disadvantaged students then grant higher marks for any given answer. Alternatively, ΔG_i may be higher in schools with disadvantaged students because teachers ask questions which are tailored to the students' knowledge or abilities. In this case, too, we consider the label "double standards" to be appropriate since it does not really matter whether grades are higher because expectations are lower, or because difficult topics are avoided.

It is, however, important to realize that whenever we speak of disadvantaged students, with the data at hand we may not be able to fully distinguish between being disadvantaged in terms of pure social status and being disadvantaged in terms of having lower ability. This is due to the fact that we cannot be certain that in the specifications presented thus far, our minority share variable measures social status net of ability. As the recommendation after primary school is a rather crude control for ability, the minority share coefficient may still be confounded with ability.

First, the advice may be a quite noisy measure of ability and therefore not fully capture all aspects of the skill distribution. Second, one could be concerned that minority and non-minority students who have received the same advice may differ systematically in their abilities due to teacher discrimination. There has been a considerable amount of research on whether recommendations after primary school may be biased towards or against students of low social status: the consensus is that while in the late 1980s and early 1990s Dutch teachers awarded higher recommendations to minority students than native students, conditional on test scores, this practice has not been observed in the time period we study (Stevens et al. 2011 provide a survey of the literature). Even though this speaks against the advice being systematically skewed towards one group of students we can still assess what effects the two possible forms of discrimination would have on the minority share coefficient: if primary school teachers correctly assess the abilities of native students and at the same time tend to underassess the true qualities of minority students and the minority students thus end up with lower secondary school recommendations, then the minority students at a given school track should be of higher ability than the native students, given a certain secondary school recommendation. In such a setting, the model predicts counteracting effects of the minority share variable: On the one hand, higher ability $\overline{\gamma}_L$ of the minority students should actually be linked to higher standards while the wage discount for low status λ_L should be associated with lower standards. In sum, the minority share coefficient would thus not capture the full magnitude of the relation between low status and low standards, i.e. it would underestimate the pure social status coefficient.

More consistent with our line of argument, primary school teachers may also issue favorable recommendations to minority students when compared to native students, given some level of ability. In that case, the minority share variable also captures the fact that minority students with the same recommendation as a native student may be of lower ability. Going back to the model terminology, this would mean that a high minority share measures a low $\overline{\gamma}_L$ – which should be associated with lower standards, too. The implication is that the coefficient on minority share most likely overestimates the pure social status coefficient, as it contains elements of ability.

To sum up, inasmuch as the recommendation after primary school is biased, we cannot fully distinguish empirically between $\overline{\gamma}_L$, which constitutes ability, and λ_L, which denotes the wage discount for low social status. In addition to the lack of empirical evidence for such biased recommendations, we should also keep in mind that the model predicts both lower ability and lower status to be associated with lower standards and so our results are consistent with the theory in either case.

A situation where students who attend minority schools receive higher recommendations is closely related to the idea of relative grading. While higher advice to minority students may be the outcome of relative recommending in primary school, a practice of relative grading in secondary education may also affect the minority share coefficient. If schools with higher minority shares have a lower average ability $\overline{\gamma}_L$, then a positive minority share coefficient may actually be the result of relative grading within school. That is, if a student is graded in relation to the ability of his peers, then students in a weaker environment in terms of ability will be awarded higher school grades, all else equal. While this explanation certainly makes sense, it is important to realize that such grading behavior constitutes nothing short of lenient grading towards students who attend ability-wise disadvantaged schools.

4.3 Robustness and reverse causality

In this section we address the possibility that the share of minority students itself may be a function of the grade difference ΔG_i, or that both variables are driven by another unobserved variable. In order to deal, first, with unobserved heterogeneity, Table 5 introduces some further specifications. In the top panel we add fixed effects for the 25 Dutch education regions defined by the Dutch Office of Statistics. Such an 'education market' is determined by student flows and characterized by offering access to all forms of higher secondary education. The education market fixed effects will thus net out any common characteristics of these markets such as local grading culture, or a clustering in space of unusually able students. Including these indicators only changes the middle track coefficient to a notable extent, while the coefficients for the other tracks remain stable.

The middle panel of Table 5 also includes the lagged grade difference ΔG_i for the respective tracks. It captures inertia in grading or historical factors that may have established a certain grading culture which may in turn be linked to minority shares. As expected, the lagged grade difference is highly significant in all cases, and it also substantially reduces the minority share coefficients in all four tracks. Importantly, however, the minority share coefficient is still statistically significant for the highest three tracks – for the lowest track there was no significant coefficient to begin with.

Finally, the bottom panel of Table 5 shows results when both fixed effects for education regions and the lagged grade difference are included in the estimations. This specification renders the minority share coefficient insignificant for the middle track, while the coefficients for the pre-university track and the higher track remain almost unchanged and statistically significant. In sum, these findings are in line with our argument that the depreciation of standards is more pronounced the higher the secondary track. Standard cutting related to disadvantaged students is still twice as large in pre-university education in terms of magnitude when compared to higher secondary education. When it comes to the lowest two tracks, it now appears that they do not differ in their standard setting behavior, which is perhaps not so surprising since the size of the relevant labor market for students in these tracks may not differ much – students from both tracks mostly move on to a similar next step, i.e., vocational training.

Table 5: ROBUSTNESS: ADDITIONAL CONTROLS, ALL SECONDARY TRACKS.

	pre-university		**higher secondary**		**middle secondary**		**lower secondary**	
add Education Region FE								
Minority students %	.109***	(0.029)	.053***	(0.018)	.026***	(0.008)	.013	(0.010)
all controls	yes		yes		yes		yes	
Education region FE	yes		yes		yes		yes	
N	635		597		802		454	
adj. R^2	.30		.24		.32		.36	
add lagged dependent variable								
Minority students %	.052***	(0.014)	.030**	(0.013)	.017***	(0.006)	.005	(0.007)
lagged ΔG_i	.537***	(0.034)	.523***	(0.112)	.623***	(0.031)	.424***	(0.040)
all controls	yes		yes		yes		yes	
Education region FE	no		no		no		no	
N	634		592		779		445	
adj. R^2	.50		.48		.54		.46	
add Education Reg. FE & lagged dep								
Minority students %	.056***	(0.016)	.027**	(0.013)	.007	(0.006)	.005	(0.008)
lagged ΔG_i	.509***	(0.035)	.514***	(0.108)	.596***	(0.033)	.374***	(0.043)
all controls	yes		yes		yes		yes	
Education region FE	yes		yes		yes		yes	
N	634		592		779		445	
adj. R^2	.51		.49		.55		.46	

Columns show estimates for the school track sample indicated in the column header. Specifications add Education Regions fixed effects and a lagged dependent variable. Education Regions are 25 'education markets', as defined by the Dutch Office of Statistics. 'All controls' are those included in Table 4 and column (3) of Table 3, respectively. The dependent variable is the difference between the school grade and the central exam grade. Standard errors in parentheses allow for clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A second possible source of concern might be reverse causality. Such a problem would arise if minority students were to choose schools with more lenient grading while non-minority students do not behave in this manner. Given the absence of catchment areas in the Netherlands, this problem could even be aggravated since students and their parents do not have to move to the vicinity of the desired school.

Intuitively, we have little reason to believe that parents of lower social status care more about their children's grades than their well-off counterparts. Quite the contrary, one would probably expect parents of higher social status to be rather more career-oriented. On top of that, it should be noted that even in the absence of catchment areas, sending an offspring to farther away schools which award better grades entails travel costs and is thus more easily feasible for well-off families.

Although this suggests that, on theoretical grounds, reverse causality is not likely to occur in our setting, we propose two ways of dealing with the issue econometrically. The first is an instrumental variables strategy that takes advantage of the fact that many institutions offer multiple school tracks, whereas the second one makes use of the fact that if students and parents choose schools according to local grading, they will condition their choice on grading policy during the previous year.

Whenever one institution offers more than one track, we expect minority shares to be highly correlated across tracks, because mostly the various tracks will be offered in the same municipality. We thus employ as instruments the minority share in all school branches but the one under consideration, leaving three instruments per school branch (e.g. the minority share in pre-university schools is instrumented with the minority shares in higher secondary, middle secondary, and lower secondary education). The shares of minority students in other school branches are obviously only available if a certain institution offers more than one track. This is not always the case, especially when the instrument is not from a school branch adjacent to the instrumented one, as can be seen from the varying number of observations in Table 6.

While the minority share in e.g. the pre-university track should not be directly driven by the grade difference in the lowest track – we do not expect pre-university track students to let the lowest track grade difference guide their choice of school – the instrument validity hinges on another assumption: clearly, the minority share in one school track is exogenous to the grade difference in another branch only if the institution does not apply an institution-wide common grading policy towards minorities, regardless of track. Since local grading policies usually emerge from the interaction among teachers and between teachers and students, they are likely to be school branch specific. Though we have no way of proving this claim, we would also expect to find lenient grading in the lower and middle secondary tracks if there were a common policy to, say, grade minority students more leniently. The results so far present tentative evidence that such grading policies don't seem to be present in the lower two tracks and in that sense, the minority shares in the lowest two tracks probably present the more convincing instruments.

As a measure of social status we now restrict attention to the minority share as the strongest predictor of grade differences. For the sake of brevity, we do not report full results of the IV regressions. Rather, Table 6 shows the second stage coefficients of the instrumented explanatory variable "minority share" as well as the first stage coefficients of the instruments.[14] In this table, we also restrict attention to the pre-university (top panel) and higher secondary education (bottom panel) tracks for which we have so far found a statistically significant relation between minority share and grading standard.

For the pre-university track we find that the coefficients are very similar to the baseline results when the higher or middle track minority share is used as an instrument. Using the lower track minority share as an instrument produces a larger coefficient. However, the difference in coefficients when compared to the OLS specification is not statistically significant. Things are similar for higher secondary education: For the first two instruments (pre-university and middle secondary education) we find a statistically significant coefficient that is very close to the OLS results. While imprecisely estimated when the

[14]Full results are available from the authors upon request.

Table 6: IV ESTIMATES, INSTRUMENTS INDICATED IN COLUMN HEADERS

	OLS	**IV: minority % pre-univ.**	**IV: minority % higher**	**IV: minority % middle**	**IV: minority % lower**
pre-university education					
minority students %	0.056***		0.065***	0.051***	0.099*
	(.016)		(0.013)	(0.015)	(0.058)
N	634		577	476	207
instrument coeff.			.728***	.512***	.227***
first stage			(0.055)	(0.081)	(0.054)
higher secondary education					
minority students %	0.027**	0.027**		0.026*	0.025
	(0.013)	(0.012)		(0.015)	(0.025)
N	592	579		487	216
instrument coeff.		0.935***		0.675***	0.314***
first stage		(0.092)		(0.060)	(0.055)

Controls are as in the bottom panel of Table 5 and include education market fixed effects and lagged grade differences. The table shows 2nd stage coefficients of the respective minority share variables, as well as the 1st stage coefficients of the instruments. Instruments are minority shares in adjacent school tracks, and are indicated in the column headers. For comparison, OLS estimates are shown in the first column. The dependent variable is the difference between the school grade and the central exam grade. Standard errors in parentheses allow for clustering at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

lowest track minority share is used as an instrument, the coefficient is again virtually the same as before.

The second way of dealing with reverse causality is to consider the fact that if parents and students were to pick their secondary school according to observed grade differences they would have to rely on grade differences from the past school year, i.e., the lagged dependent variable. This variable is already included in our estimations and it is worth stressing that even if one were to doubt the instrumental variable validity, the OLS estimations which include the lagged grade difference also account for reverse causality issues – the fact that minority share remains significant in these specifications makes a strong case against reverse causality driving the results. In the sense that these specifications do not rely on an untestable assumption about the instrument, they are possibly more convincing than the IV. Taken together, the results we found in this section therefore do not suggest that either unobserved heterogeneity or reverse causality drive the earlier results.

5 Conclusion

In this paper, we analyze the impact of social class and ability on the choice of grading standards by schools. We show in a theoretical model that schools with a student body that is disadvantaged in terms of either pure status or ability tend to apply less demanding standards. The predictions of the model are then tested on data from the Netherlands since the Dutch educational setup provides the rare opportunity of measuring decentralized grades awarded by the individual schools against the benchmark of central test results. In accordance with the theory, the empirical results show that schools with many students from cultural minorities, or with many students receiving financial aid, award higher school grades.

It is worthwhile to compare this result with empirical findings for Germany. Based on the PISA study, Prenzel et al. (2005) claim that students from lower classes are discriminated against in the grading and examination system. Similarly, more recent research shows that in primary education, pupils with an immigrant or otherwise disadvantaged background obtain lower grades (Kiss, 2010) and lower track recommendations[15] (Lüdemann and Schwerdt, 2010) than others, even after controlling for ability as measured by the PIRLS test. In contrast our results suggest that such students are held to less demanding standards than students from average backgrounds.

There are several possible explanations for this apparent difference. Obviously, the disparity in grading standards applied may simply reflect different attitudes towards immigrants and lower classes in the two countries studied. More subtly, school financing may play a role. Schools in the Netherlands obtain a higher per capita funding for disadvantaged students, so that they have a larger incentive than German schools to treat such students favorably.

[15] According to Jürges and Schneider (2011), also boys and younger students obtain lower track recommendations.

A second kind of explanation may be provided by the different nature of the data used. We examine school aggregates, whereas Kiss (2010) and Lüdemann and Schwerdt (2010) compare immigrants and natives inside a class. Thus, it might be possible that in both countries, classes with a high share of immigrant or socially disadvantaged students are held to lower standards, but that inside the class, the native and socially favored students benefit more from these lower standards. From a theoretical point of view, however, school specific standards appear more convincing than discrimination inside classes. As our model shows, lenient grading by a disadvantaged school increases the wages earned by this school's students, which in turn may benefit the school, for example by raising its reputation among parents and so attracting future students. Contrary to that, it is unclear what motivation might drive teachers to discriminate against individual students.

A third explanation refers to the fact that the benchmarks against which the grades are measured serve different purposes. The PIRLS and PISA tests used in the German studies measure general mathematical or language abilities, which certainly are highly correlated with exam results, but are not tailored specifically to the material covered at school. In contrast, the central examination in the Netherlands is designed specifically to test knowledge of the schools' curriculum, just as the local examinations. Thus, in the German case, a divergence of grades from test results may to a larger extent than in the Netherlands be driven by the divergence in aim and nature of the benchmark used from the aim and nature of school examinations.

Finally, our results pertain to secondary education while the results on Germany cited above relate to primary education. This may be important since in primary education, grading is likely to be influenced in a much stronger way by a teacher's personal impression of the individual student than in secondary education. Consequently, a personal bias which the teacher may harbor against immigrants and other disadvantaged groups may more easily influence the grades in primary than in secondary schools. Evidence supporting the view that the level of schooling matters for the grading policy is provided by Kiss (2010). Unlike in primary education, Kiss does not find a grade disadvantage for immigrants in secondary schools, and in some specifications he even, quite in line with our results, finds that these students obtain better grades.

Taken together, it is however difficult to decide which of these explanations drives the results. Thus, we refrain from drawing any more general conclusions from our findings at this stage. Nevertheless it seems safe to point out that, in order to explain the treatment of lower class students by the schooling system, more than a simple appeal to discrimination is needed.

This observation suggests that much more research is required in order to enhance the understanding of how standards are set. For example, it will be fruitful to integrate other motives for the choice of standards. As some of our empirical results suggest, competition for students may be an important driver of grade inflation. It is worthwhile to analyze this, both theoretically and empirically, in more detail in future work.

References

Betts, J. (1998a), 'The Impact of Educational Standards on the Level and Distribution of Earnings', *American Economic Review* 88, 266-275.

Betts, J. (1998b), 'The Two-legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools', *Economic Policy Review* 4, 97-116.

Betts, J. and R. Costrell (2001), 'Incentives and Equity Under Standards-based Reform', *Brookings Papers on Education Policy*, 9-74.

Bishop, J. (1997), 'The Effect of National Standards and Curriculum-based Exams on Achievement', *American Economic Review* 87, Papers and Proceedings, 260-264.

Bishop, J. (1999), 'Are National Exit Examinations Important for Educational Efficiency?', *Swedish Economic Policy Review* 6, 349-398.

Bishop, J., F. Mane, M. Bishop, and J. Moriarty (2001), 'The Role of End-of-course Exams and Minimum Competency Exams in Standards-based Reforms', *Brookings Papers on Education Policy*, 267-345.

Burgess, S. and E. Greaves (2009), 'Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities', *CMPO Working Paper* No. 09/221, University of Bristol.

Centraal Bureau voor de Statistiek, Kerncijfers postcodegebieden 2003.

Centraal Bureau voor de Statistiek, Kerncijfers wijken en buurten 2001-2005.

Chan, W., L. Hao, and W. Suen (2007), 'A Signalling Theory of Grade Inflation', *International Economic Review* 48, 1065-1090.

Costrell, R. (1994), 'A Simple Model of Educational Standards', *American Economic Review* 84, 956-971.

Costrell, R. (1997), 'Can Centralized Educational Standards Raise Welfare?', *Journal of Public Economics* 65, 271-293.

Dronkers, J. (1999), 'Is het eindexamen wel gelijkwaardig tussen scholen? Discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen in het voortgezet onderwijs.' *Farewell address from the chair of educational sciences of the University of Amsterdam.*

Entorf, H. and N. Minoiu (2005), 'What a Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration', *German Economic Review* 6, 355-376.

Epple, D., E. Newlon, and R. Romano (2002), 'Ability Tracking, School Competition, and the Distribution of Educational Benefits', *Journal of Public Economics* 83, 1-48.

Epple, D. and R. Romano (1998), 'Competition Between Private and Public Schools, Vouchers, and Peer Group Effects', *American Economic Review* 88, 33-62.

Hanushek, E. (2002), 'Publicly Provided Education', in: A. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics* 4, Elsevier, 2047-2141, Amsterdam.

Inspectie van het Onderwijs, Kwaliteitskaart Voortgezet Onderwijs, uitgaven, Steinmetz Archief, Amsterdam.

Jacob, B. (2005), 'Accountability, Incentives and Behavior: The Impact of High-stakes Testing in the Chicago Public Schools', *Journal of Public Economics* 89, 761-796.

Jürges, H., W. Richter, and K. Schneider (2005), 'Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality', *Finanzarchiv N.F.* 61, 298-326.

Jürges, H. and K. Schneider (2011), 'Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System', *German Economic Review* 12, 371-394.

Jürges, H., K. Schneider, and F. Büchel (2005), 'The Effect of Central Exit Examinations on Student Achievement: Quasi-experimental Evidence From TIMSS Germany', *Journal of the European Economic Association* 3, 1134-1155.

Kee, P. (1995), 'Native-Immigrant Wage Differentials in the Netherlands: Discrimination?', *Oxford Economic Papers* 47, 302-317.

Kiss, D. (2010), 'Are Immigrants Graded Worse in Primary and Secondary Education? Evidence for German Schools', *Ruhr Economic Paper* 223.

Lüdemann, E. and G. Schwerdt (2010), 'Migration Background and Educational Tracking: Is there a Double Disadvantage for Second-Generation Immigrants?', *CESifo Working Paper* 3256.

Ministerie van Onderwijs Cultuur en Wetenschappen, Onderwijs in Cijfers 2003.

Nechyba, T. (1999), 'School Finance Induced Migration and Stratification Patterns: The Impact of Private School Vouchers', *Journal of Public Economic Theory* 1, 5-50.

Prenzel, M., J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, and U. Schiefele (eds.) (2005), *PISA 2003: Ergebnisse des zweiten internationalen Vergleichs – Zusammenfassung*, IPN - Leibniz-Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel, available under http://pisa.ipn.uni-kiel.de/Ergebnisse_PISA_2003.pdf.

Sociaal en Cultureel Planbureau, Statusscores postcodegebieden 2002.

Stevens, P., N. Clycq, C. Timmerman, and M. Van Houtte (2011), 'Researching race/ethnicity and educational inequality in the Netherlands: a critical review of the research literature between 1980 and 2008', *British Educational Research Journal*, 37(1), 5-43.

van Ours, J. and J. Veenman (2004), 'From Parent to Child: Early Labor Market Experiences of Second-Generation Immigrants in the Netherlands', *De Economist* 152, 473-490.

Wikström, C. and M. Wikström (2005), 'Grade Inflation and School Competition: An Empirical Analysis Based on the Swedish Upper Secondary Schools', *Economics of Education Review* 24, 309-322.

Wößmann, L. (2003), 'Schooling Resources, Educational Institutions and Student Performance', *Oxford Bulletin of Economics and Statistics* 65, 117-170.

Zorlu, A. (2002), 'Ethnic and Gender Wage Differentials', *University of Amsterdam AIAS Research Report* 13.