Administrative groupings and equity in public service provision

Sarah Yde Junge[1, 2, 3]

[1] *Department of Political Science, Aarhus University, Aarhus C, Denmark*

[2] *TrygFonden's Centre for Child Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark*

[3] *Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark*

2020

November

Abstract

Scarcity in public service agencies requires that available resources are prioritized for citizens who need them most, and inherent to all prioritizations is a comparison of the cases. Despite the amount of research that has been conducted on the prioritization process, surprisingly little attention has been paid to the underlying comparison. This study suggests that the administrative grouping of citizens in public service agencies influences the comparison so that the allocation of public services not only depends on a citizen's need, but also on the needs of the other citizens assigned to the same agency. Based on registry data on more than 300,000 students, this study exploits the as-good-as-random assignment of students to classes within

schools to analyze the influence of classroom composition on referrals to special education in Danish primary school. The result shows that students with better-performing peers are considerably more likely to be referred to special education than equally low-achieving students in lower-performing classes. This finding illustrates an unwarranted distributive consequence of the administrative grouping of citizens, and thereby initiates a discussion of how groupings affect service provision and how such an influence can be mitigated.

Introduction

Citizens receiving public services are almost always grouped in some way in order for the bureaucracy to be able to provide them the service. Examples are students in schools and classes, toddlers in preschools, patients in hospital wards, prisoners in blocks, the assignment of patients to a general practitioner, the group of clients a specific caseworker handles, the assignment of litigation to judges or the division of neighborhoods in police districts. These administrative groupings of citizens may seem simply to be an organizational arrangement of cases, but the question is whether the grouping in itself also influences who gets which services, when and how. There is a long tradition of studies of differential treatment of citizens in street-level bureaucracies (e.g. Andersen and Guul, 2019; Harrits and Møller, 2014; Soss, Fording, and Schram, 2011b and see Tummers et al., 2015 for a review). This study investigates a new source of differential treatment, which—contrary to previous research—is not rooted in characteristics of the street-level bureaucrat or the citizen, but in the grouping of citizens within an agency. Harrits (2019) argues that the heterogeneity of the group of citizens in a specific public agency may affect street-level bureaucrats' use of stereotypes, and Lipsky suggested in his groundbreaking work *Street-level Bureaucracy* that the group of citizens in a given public service may affect which of the individual citizens is determined eligible for the given service (Lipsky, [1980] 2010). Despite the amount of work on street-level

bureaucracy and despite the obvious unfairness of the administrative grouping of citizens as a determinant of the service level provided to each individual citizen, the consequences of the grouping itself have not yet received much attention in the public administration literature.

The aim of this study is to combine different theoretical insights in a theory of how the administrative grouping of citizens in public services affects the service provided for each citizen, and thereafter to test this theory. Legal regulation as well as the bureaucratic organization of public service agencies ought to secure equal treatment of citizens (Weber, 1947). However, theoretically, one could provide at least three arguments for why the grouping of citizens could in fact influence the service provided to each individual citizen. First, the other group members may claim a smaller or larger part of the available resources and thereby influence what is left for any individual group member. Second, when street-level bureaucrats assess each citizen's need for a given service, the needs of other group members may be used as an anchor; that is, the baseline for comparison (Lipsky, [1980] 2010; Tversky & Kahneman, 1974). Last, the group composition may signal something to the individual group member about her own needs and abilities that could make her change her demands for public service (Festinger, 1954). While the first argument may be a legitimate result of different political prioritizations, the other two reasons to expect an influence from the administrative groupings are not. This study will keep the first argument fixed, and investigate whether one citizen's eligibility for public services depends on the needs of the other citizens who belong to the same administrative group beyond the differences in available resources.

Administrative groups may be difficult or unethical to experimentally manipulate, and non-manipulated groups are very often non-randomly created, which may cause studies of these groupings to be biased. Therefore, the primary concern in studies of administrative groupings is to overcome selection issues in a non-experimental setting. To explore the influence of grouping and at the same time overcome selection issues, this study relies on the grouping of students in classrooms within schools. A fixed effects design enables me to deal

3

with selection into schools and uneven allocation of resources between schools, and formal tests of assignment to classes show that classroom formation within schools is balanced on all important observable variables, and thereby provides exogenous variation in the groups. I use this variation to investigate how the administrative grouping of students affects referrals to special education in Danish public primary school. The Danish registers contain population-wide data (N = 314,405) on special education referrals, classroom composition and academic performance in national standardized tests, as well as a comprehensive list of background information. Using this, I can investigate how variations between classes within schools may affect the probability of an individual being referred to special education (Ammermueller & Pischke, 2009; Hoxby, 2000). This case uniquely captures the complexity of decisions on public service provision, and thereby allows me to test the theoretical argument about the influence of grouping on public service provision.

In line with the theoretical claim, the results show that students with better-performing peers are more likely to receive special education. Additionally, referred students' need for special education is also evaluated by external experts, and their evaluations are not correlated with classroom composition. Thus, the analysis supports the theoretical argument that the administrative grouping of students affects the decision of special education referral.

This study draws attention to a new source of inequality in public service provision, which has a number of implications for both public administration research and the organization of public service provision. First, it builds a theoretical argument about how administrative groupings may cause inequality and raises a discussion about how this may interact with our current knowledge on differential treatment and how it can be mitigated. Second, it calls for further research into the influence of administrative grouping and the general relevance of this phenomenon across different street-level bureaucracies. Finally, an awareness of inequalities as a result of administrative groupings of citizens may inform the processes of decentralization and organizational design of street-level bureaucracies and increase the likelihood of public

services being allocated fairly.

Theory

Administrative grouping and prioritization

The overarching theoretical aim of this study is to combine organizational decisions about the assignment of citizens to public services with micro-level decisions on service provision. As exemplified above, citizens are almost always grouped in some way in order for the bureaucracy to be able to provide them public services. Since the grouping is a precondition for the delivery of many services, it is relevant to understand how the grouping in itself may influence service provision.

Naturally, administrative groupings will look different in different public organizations. The groups may be identifiable at different levels, as clients of a specific public agency, and also as parts of smaller units within the agency. The citizens may have a smaller or larger say in which group they are assigned to. In schools, daycares, and nursing homes, for example, citizens often hold some knowledge on the difference between institutions and they may also have quite substantial influence over which institution they (or their relatives) are assigned to, and hence which group they become part of. Lastly, the grouping may be more or less visible to the citizens. In schools, the group members interact with each other, and thereby the grouping is clear to everyone, whereas patients assigned to a specific general practitioner may not know or interact with the other patients treated by their doctor. Despite the differential nature of groups, all administrative groupings share the characteristics that they are administratively created and known by the public service agency.

The administrative groups may become important when a scarcity of resources requires prioritization among citizens. Because there is no price limit on the demand for public services, it very often exceeds the supply and necessitates a prioritization of cases (Lipsky,

5

[1980] 2010). One way to make such prioritizations is to direct services to those citizens within the agency with the greatest needs. Delivery of public services is often decentralized, which leads different agencies to have different resources available and to face differential needs from their clients. If there is no financial equalization of the differences in demand and supply, citizens in low-need groups will experience a higher service level than equally deserving citizens in a high-need grouping (Boadway & Shah, 2009). Thus, the influence of the administrative grouping may depend on the financial equalization between agencies. While decisions of financial equalization and hence on the accepted differences in service levels between agencies are politically determined, the administrative grouping of citizens within agencies should not cause differences in service levels. Therefore, the remaining part of this study will discard the prioritization argument and focus on the two other arguments for an influence of administrative grouping in the provision of services to citizens within the same agency.

According to Weber, a fundamental advantage of modern bureaucracy is that "Everyone is subject to formal equality of treatment, that is, everyone in the same empirical situation (Weber, 1947, p. 340). Within agencies, everyone draws from the same pool of resources, and therefore differential treatment of similar cases is not caused by differences in the available resources. However, differences in resources are not the only reason for differential treatment. Numerous studies have shown that an inconsistency between supply and demand may cause differential treatment of citizens within the same agencies (e.g. Andersen and Guul, 2019; Harrits and Møller, 2014; Soss, Fording, and Schram, 2011b and see Tummers et al., 2015 for a review). Below, I will outline two theoretical arguments—a supply-side and a demand-side argument—for why administrative groupings within agencies may also be a source of differential treatment of empirically similar cases.

Micro-level arguments for an influence of grouping

The two theoretical arguments for an influence of administrative groupings rely on the same logic about the administrative group as a relevant comparison for assessments of needs. The first argument is a supply-side argument concerned with how the front-line worker makes assessments of needs and eligibility, and the second argument is a demand-side argument concerned with citizens' perceptions of their own needs. Below, I will shortly elaborate the two arguments and relate them to theories and findings from related literatures. They will not be separated in the empirical analysis of this study; instead they are used as a theoretical foundation to understand why we could expect administratively created groups to influence the services provided for citizens in public agencies, beyond the politically accepted differences caused by differential resources between agencies.

The supply-side argument for why the administrative grouping of citizens may affect the public service provided to individual citizens is that the grouping may affect the assessments that street-level bureaucrats make. So far, research on street-level bureaucrats' decision making has mainly focused on the encounter between the citizen and the street-level bureaucrat, and determinants of street-level bureaucrats' behavior have been found to be linked to bureaucrats' views of themselves, but also to characteristics and perceptions of the clientèle (e.g. Jilke and Tummers, 2018; Keiser, 2010; Maynard-Moody and Musheno, 2003; Soss, Fording, and Schram, 2011a). Harrits (2019) finds indications that street-level bureaucrats' use of stereotypes is affected by the composition of citizens assigned to their agency. Thus, front-line workers who work in a heterogeneous social context are less likely to use stereotypes in their encounters with citizens than front-line workers working in a more homogeneous social context. Likewise, Einstein and Glick (2017) find that racial disparities in the tone of replies on social housing requests were smaller in agencies with a more heterogeneous clientèle. These findings illustrate that the evaluation and treatment of citizens in public service agencies may be influenced by the group of citizens the street-level bureaucrat works with.

An argument for such a correlation could be found in psychological work on human processing of information and construction of judgment, which has repeatedly demonstrated that humans cannot make absolute judgments. Instead, they use existing knowledge or prior experiences to anchor their assessments. Thus, human evaluations always rely on some kind of comparison (Helson, 1947; Kahneman, 1992; Tversky & Kahneman, 1974).

In public administration research it has been demonstrated that managers use prior performance and the performance of neighboring organizations as reference points that inform expectations and aspiration levels (Holm, 2017; Nielsen, 2014; Simon, 1939), and other studies have investigated the extent to which citizens also rely on comparisons when they evaluate performance information on public services (Andersen & Hjortskov, 2016; Olsen, 2017). However, whether street-level workers compare their clients to determine eligibility is still unknown. In his work on street-level bureaucracy, Lipsky mentioned in passing that *"it is probably fair to say that clients will always be differentiated in terms of their perceived relative normality, regardless of how absolutely receptive to intervention they are"* (Lipsky, [1980] 2010, p. 113). Thus, Lipsky suggested that street-level bureaucrats also rely on some kind of comparison to other clients when they evaluate clients' needs. Hence, the argument is that the administratively created group of citizens that a street-level bureaucrat is working with affects her perception of what is normal, who is in need of public services, and who is considered eligible for them.

Though the comparison of clients may not have received much attention in itself, one could argue that it is already an implicit, underlying assumption in much work on street-level bureaucracy and the coping mechanisms employed by front-line workers. One example is *creaming* or *prioritizing*. To prioritize the easiest clients, street-level bureaucrats compare their clients to each other to figure out who they can most easily succeed with, and thereby who they should care most about in order to reach their performance goals. These types of behavior necessitate a comparison of cases, but the comparison in itself has not received

much attention in public administration research.

Beyond public administration, some empirical studies have suggested that the grouping of citizens affects street-level bureaucrats' assessments. For example, studies of grading and teachers' recommendations of further education for their students suggest that teachers use the classroom as an anchor or a frame of reference, which implies that students get better grades (Calsamiglia & Loviglio, 2016; Møen & Tjelta, 2010; Neumann, Trautwein, & Nagy, 2011), and are more likely to get better recommendations for further education, when their peers do worse (Boone et al., 2018). However, most of these studies do not methodologically address the selection into groups, and they all study a quite unidimensional assessment, with no economic costs associated. Public services are often based on more complex assessments, and the decision to provide a service for a citizen is associated with a cost. The studies above inspire the argument that administrative groupings may affect service provision, but so far there is no empirical evidence that their conclusions will withstand a harder test and travel to more complex cases like provision of services.

The demand-side argument instead points to the citizens. The argument is that the grouping itself may signal something to group members about their own status, which makes them change their demands for public service. This argument emanates from the social comparison theory, first introduced by Festinger (1954). The basic claim is that humans rely on reference groups when they evaluate their own abilities and opinions, and may think more or less of themselves depending on the reference group. Empirical studies have demonstrated the importance of the reference group in assessments of academic self-efficacy, showing that students evaluate their own performance more favorably and have higher aspirations when they perform relatively better than their peers (Davis, 1966; Marsh, 1984; Marsh et al., 2008; Seaton, Marsh, & Craven, 2010). These studies indicate that humans' assessments of themselves adapt to their surroundings. One could imagine that this adaptation would also occur in public services, such that citizens grouped with citizens with low needs would think

more of their own needs and hence increase their demands for public support.

Together these two arguments form the hypothesis that *a citizen in an administrative group where the other group members have lower service needs is more likely to receive public services than the same citizen would be in an administrative group with higher service needs.*

Empirical design

There is one major challenge to studying the consequences of administrative groupings, namely to overcome the selection issues. Even when citizens do not decide who they will be grouped with, administrative groups are seldom created completely at random, and therefore the consequences of groupings are hard to separate from the constituent reasons. It is rarely possible for researchers to experimentally manipulate groupings, and the complexity of a group may be hard to convincingly simulate in a survey-experimental design. Therefore, the study of groupings needs to rely on observational data with some kind of exogenous variation in the group formation.

I use register data on special education referrals in Danish primary schools to examine the consequences of administrative groupings for service provision. Formal tests of differences between classes on predetermined student characteristics show that assignment to classes within schools is as good as random. These differences between classes within schools provide exogenous variation to the administrative groups under study, which enables me to overcome selection issues. Furthermore, I rely on school and register data, which besides providing a sample large enough to detect even smaller effects also gives me very high data validity. Below, I will introduce the case of special education in the Danish school system and the estimation strategy applied, followed by an operationalization of the measures used in the study and a discussion about the assumption of random assignment to classes.

Special education referrals in Danish schools

In Denmark, there is a voucher system that allows Danish parents to choose which school to send their children to. There is thus considerable selection into schools, but the assignment to classes within schools does not seem to suffer from this selection. Assignment to classes is done at the school level. National regulations of class formations state that there can be a maximum of 30 students per class and that the local governments must make guidelines for class formation (UVM, 2017). Most municipalities only make guidelines about class sizes, and let the schools themselves decide how to assign students to classes. A look into different school policies on class formation shows that most schools aim for a balanced gender distribution, and some also actively assign students to classes in ways that secure a social and geographic balance between classes. In some schools, parents may name a child who they want their child to be assigned to class with. The school principal assigns teachers to classes. Each class has one or two head teachers, who typically teach Danish and math. Head teachers are normally assigned to a class for a 3-5 year period, so the head teachers in 1st grade will follow the class until 3rd to 5th grade.

When there are students in a class whose "development necessitates special consideration or support that cannot be given within the setting of the ordinary class" (UVM, 2010), they are eligible for special education. Students can be referred to special education for many different reasons, but the purpose of special education is for the students to acquire knowledge and competencies that will prepare them for further education. Therefore, academic challenges need to be present no matter what other difficulties the student may struggle with (UVM, 2010, 2017).

Detailed student-level data on special needs are rarely available, and especially not for the entire population of students, but in the school year 2011-12, special education referrals and the magnitude of special education was registered for every student in the Danish school system. Performance data on an individual level is available from 2nd grade for all public

school students, and therefore I can include everyone in ordinary classes in public schools in grades 3 to 9 in the school year 2011-12 in the analyses. This gives me a sample of 314,405 students. This unique data source is the basis of this study.

While special education can be organized in several different ways, the most important distinction for this study is whether the student with special needs is enrolled in an ordinary class or in a special needs class. Only students in ordinary classes are included in this study. Students in this category will attend an ordinary class most or all of the time and then have either an extra teacher to support them during some of the lessons or have some individually scheduled lessons in addition to or instead of ordinary class. Even though there is some debate about the effect of special education (Hanushek, Kain, & Rivkin, 2002; Keslair, Maurin, & McNally, 2012; Kvande et al., 2019), there is no doubt that special education is given to students in an attempt to support them. This is underlined by the amount of money spent on special education. A report on special education in Danish primary schools estimates that the average costs of a student referred to ordinary special education is 168 % of the costs of an ordinary student (Ministry of Finance, 2010).

When teachers feel that a student might benefit from special education, they are obliged to recommend the student for such. The recommendation is received by the headmaster or a committee at the school that takes care of special education. From here, the standard procedure is that the headmaster or the committee (in consultation with the parents) will ask a centrally organized pedagogical-psychological advisory unit (PPA unit) to make an assessment of the student. If the PPA unit finds that the student should be granted special education, it will also make a recommendation about the type and amount of special education the student should be granted. Thus, the teachers do not decide who will receive special education or how much; but if the teachers are not aware of a student's challenges or needs, the process will not be initiated. Hence, the teachers are important for the decision of who will receive special education, whereas the PPA units determine both who will receive special

education and how much. When a student has been referred to special education, her needs and the type of support granted have to be reconsidered at least once a year.

The frequency of special education referrals differs greatly between Danish municipalities. A governmental report from 2010 stated that "It is assessed that the variations between the municipal referral patterns are an expression of the fact that there is no clear and uniform practice across the municipalities as to what types and degrees of issues are dealt with in ordinary special education" (Ministry of Finance, 2010). The lack of a clear and uniform practice implies that special education referral is a service that is provided to students on a heavily discretionary basis.

Estimation strategy

As argued above, a simple comparison between schools would not necessarily shed light on the impact of groupings in schools, because the selection into schools may cause the student populations to differ between schools. To overcome these selection issues, this study applies a fixed effects design. Inspired by Hoxby's (2000) exploitation of idiosyncratic cohort-to-cohort variations in her study of peer effects, I examine whether class-to-class differences within the same school are systematically associated with referrals to special education.

The main specification of the model applied to investigate the influence of the group on special education referrals is:

$$y_{ics} = \alpha_s + \beta \overline{X}_{cs} + \gamma X_{ics} + \delta \mathbf{Z}_{ics} + \eta \overline{\mathbf{Z}}_{cs-i} + C_{cs} + \mu_{cs} + \epsilon_{ics}{}^{[1]},$$

where y_{ics} is an indicator of whether student i in class c in school s is referred to special education. $\overline{X}_{cs}$ is the grouping effect that is the average class need for special education.

---

[1] In line with the theoretical argument about the teacher's general impression of the class, the analyses use the simple average of the class' needs and the more appropriate leave-me-out average of the background characteristics. The alternative specification with a leave-me-out average of the class' needs does not change the results significantly.

X_{ics} is the student's individual need for special education. α_s is the school fixed effects. $\mathbf{Z}_{ics}$ is a vector of individual background characteristics, $\overline{\mathbf{Z}}_{cs-i}$ is a vector of leave-me-out class averages on background characteristics and C_{cs} is an indicator of the grade level. μ_{cs} is the class-level error term and ϵ_{ics} is the individual-level error term. μ_{cs} represents the correlated effect, which is the unmodeled class-specific influences. If μ_{cs} is correlated with the class composition, this will cause a bias in the estimates. If students and teachers are randomly assigned to classrooms, the correlated effect will not bias the model. The school fixed effect α_s removes any selection effect caused by selection into schools by removing all variation between schools, including any differences in available resources.

As a supplement to the analysis of referrals, I conduct an analysis with the amount of special education each student receives as the dependent variable. As described above, the amount of special education is externally determined, and hence the classroom is not expected to influence the amount of special education that referred students receive. If the amount of special education is correlated with the composition of the class, the classroom measure may be correlated with critical omitted variables. If the amount of special education is in fact independent of the classroom composition, a dependency between classroom composition and referrals are very likely a result of the influence from the administrative grouping of students.

Given the dichotomous nature of the dependent variable, I estimate the effect of classroom composition using a linear probability model. The linear probability model has the advantage of being more efficient and easier to interpret than its nonlinear counterparts (Angrist & Pischke, 2008, p. 107), especially in hierarchical models. Robust standard errors are applied in all specifications of the model to account for the heteroskedasticity imposed by the OLS estimator when estimating models with a binary dependent variable. The main specification is also estimated using a constrained conditional logistic regression. Following Beck (2020), the results from the logistic regression as well as the results from a linear regression using the logistic sample are available in the appendix. The effect sizes from these estimations are

substantial and statistically significant like the results from the main specification presented below,

Measures

A crucial part of the concept of administrative groupings is to define which characteristics of the group may influence the service provided for each individual group member. The influential characteristics are probably a multi-faceted pool; however, given the proposed mechanisms above, I will argue that the relevant characteristics of other group members need to be the characteristics that determine their needs or eligibility for public services. No matter whether I consider the supply-side or the demand-side argument, the relevant characteristics of the other group members are their need for service. Special education referrals are complex decisions and one may assume that many different factors could influence students' needs for special education. However, since academic performance is of central concern to decisions about special education, this study uses the academic performance of the class as the group measure, despite the fact that actual assignment to special education probably relies on more nuanced assessment.

I rely on the Danish national tests in reading to measure the academic ability of the individual student and the class. The national tests are computer-based, adaptive and self-correcting (Beuchert & Nandrup, 2014). This means that human bias in their correction is minimized and, even more important, the students' teacher has no influence on the scoring of the test. The tests are mandatory for all public school students in 2nd grade, 4th grade, 6th grade and 8th grade.

The national tests are taken in the spring, but special education is registered in October. To make sure that the test scores are measured prior to special education referral, I use the test scores achieved in the previous year for students in grades 3, 5, 7, and 9, and two years earlier for students in grades 4, 6, and 8. The test scores are standardized within

grade and year on the population of test scores. The class measure is the average test score performance of the class. It seems unlikely that teachers actually calculate this, but the teachers are assumed to have an impression of the general academic level in the classroom, and I expect the average performance of the class to mirror this impression.

The dependent variable of the study is the dichotomous variable indicating whether a student is referred to special education. It takes the value 1 if the student is referred to special education and 0 otherwise. 7.23 % of the students in the sample were referred to special education. An alternative specification of the model uses the hours of special education a student receives per week. For students not referred to special education, this variable takes the value 0. The information on the amount of special education is known for 77.3 % of the referred students. On average they received 1 hour and 36 minutes per week, though the average amount of special education received by special needs students may be lower since the 22.7 % whose amount of special education is unknown performed slightly better academically (mean -0.77, sd 1.03) than the 77.3 % who were observed (mean -0.86, sd 0.99)

A set of background variables obtained from the Danish registries is included in the analyses. These are gender (0=girl, 1=boy), immigrant status (0=Danish, 1=immigrant or descendant of immigrants), psychiatric diagnoses (0=no diagnosis, 1=diagnosis) and average parental education in years. In cases where only one parent's education is known, the measure relies on this parent alone. These measures are included as control variables at the individual level as well as classroom leave-me-out averages. As a measure of resource allocation, class size as registered at the beginning of the school year is included in a supplementary analysis.

Because the administrative groups are a crucial component of this analysis, one needs to think carefully about how missing data on the individual level may affect the group measures. To be sure of the reliability of the measures of class composition, classes where more than three students have not taken the test are excluded. Analyses of the subsample of cohorts with no missing test scores (see Table A1 in the appendix) reveal an effect size larger than

16

Table 1: Descriptive statistics

Number of schools	1330
Number of classes	15750
Classes per school cohort	3.178 (1.1391)
Class size	21.461 (3.596)
Special education	0.072
Test score	0.041 (0.959)
Boy	0.502
Immigrants	0.097
Psychiatric diagnosis	0.066
Parental education	14.220 (2.091)

the one reported here. This implies that the estimates reported in the results section are not driven by uncertainties of the true class average. Furthermore, observations in classes with fewer than five students or more than 30 students are excluded, since these classes are likely to be a result of an error in the registration process rather than the real class size. All class-level variables are standardized based on the sample of classes used in this study.
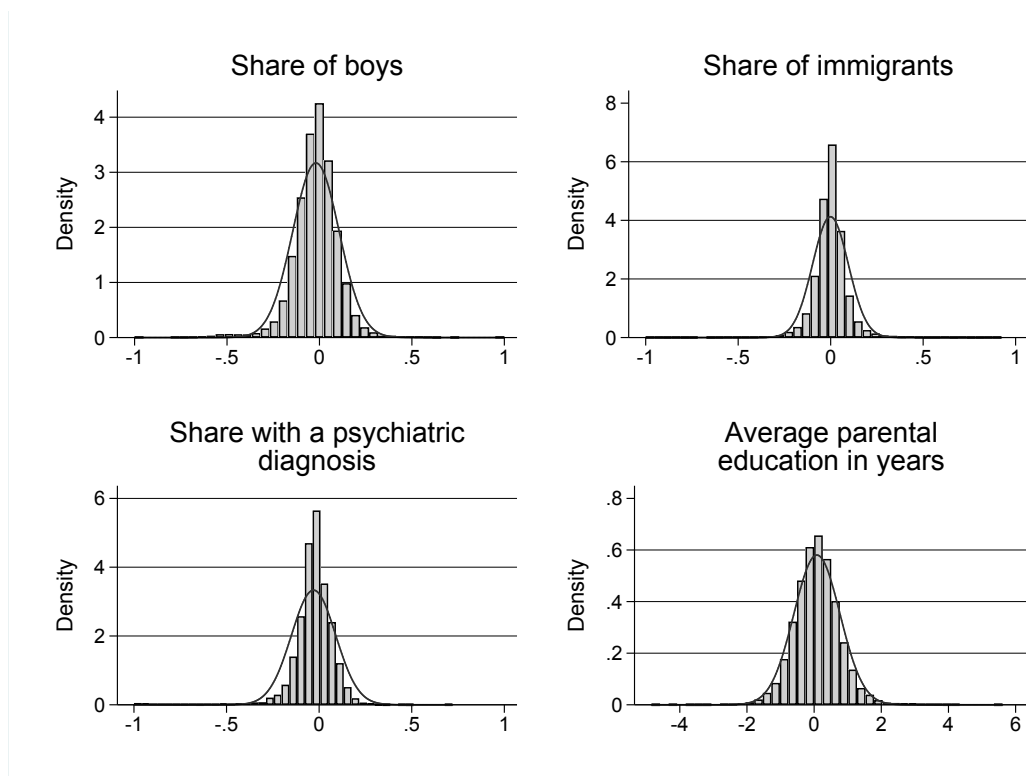
Assignment to classes

One could worry that parents not only select into schools, but also into classes within schools. There is nothing in the guidelines on class formation that indicates that this is the case, but they do not exclude the possibility that parents may interfere in class formation. To be sure class formation is not subject to parental selection, I conduct a series of tests below to formally test the assumption of random assignment to classes within schools. Figure 1 shows histograms of the deviation from the classroom average to the average of the remaining cohort within a given school on predetermined measures. If differences between classes are random, the deviations pictured in Figure 1 would be normally distributed. As can be seen in

17

Figure 1, the deviations for all variables follow the plotted normal distribution curve, though the differences seem to be a bit smaller than one would expect from completely random assignment. Following the procedure used by Ammermueller and Pischke (2009), I also make a statistical test of whether students are assigned to classes in any systematic way using a Pearson χ^2 test for observable characteristics that could potentially cause selection bias. For parental education, which is a continuous variable, I use the Kruskal-Wallis test. The p-values from these tests are plotted in Figure A1 in the appendix. If students are randomly assigned to classes, the p-values would be evenly distributed. A right-skewed distribution indicates less difference than would be expected from random allocation, whereas a left-skewed distribution indicates bigger differences than a completely random allocation would create (Ammermueller & Pischke, 2009). As also seen in the plot in Figure 1, gender is highly left-skewed. This distribution is expected, because many schools deliberately try to balance gender across classes. The p-values for parental education, psychiatric diagnoses and immigrants are close to evenly distributed, which is consistent with a random assignment of students to classrooms.

The principles of class formation imply that the patterns revealed by the Pearson χ^2 and Kruskal-Wallis tests may not be an expression of true random assignment, but rather a deliberately balanced assignment, at least for some of the variables. However, since the allocation is balanced, the variation in peer composition across classes can be considered as good as random for statistical purposes (Calsamiglia & Loviglio, 2016). Furthermore, I include the background characteristics as control variables in the models, to account for any non-randomness in these. Taken together, the tests above indicate that the underlying assumption about as-good-as-random assignment to classes within schools is not violated, and hence that the fixed effects approach manages to overcome selection issues. Another concern from μ_{cs} is that resources and teachers may be assigned to classes dependent on the classroom composition. This concern is harder to deal with, because there is no information

Figure 1: Classroom deviation from school-cohort average

on teacher assignment available. Ammermueller and Pischke (2009) argue that "teacher-shopping"—that is, teachers changing classes due to the ability of the students—is not very common in European school systems, which would imply that the teacher assignment is not problematic in this case. To try to overcome the resource question, I conduct two robustness checks. First, I investigate whether a smaller class size is correlated with fewer referrals. If the classes are smaller, one would expect the teacher to have more time within the ordinary class to support students with special needs. Second, I limit the analysis to 3rd grade students, the earliest year for which data is available. Teachers in Danish primary school are typically assigned to classes for a three- to five-year period, so most students will have the same teacher in Danish from grades 1 through 3. When the students start, the school may not know which of the classes will benefit the most from a better teacher, and hence the assignment of teachers in the lower grades may be considered more random than in higher grades.

A final concern to be mentioned here is whether there is sufficient variation between classes within schools to estimate the effect of classroom composition on special education referrals. A variance decomposition of class-level averages of reading scores is shown in Table 2. Not surprisingly, the majority of the variation in class-level reading scores is between schools (around 70 %). However, there is still substantial variation in class-level average left within schools to undertake the analyses in this study (Ammermueller & Pischke, 2009)

Table 2: Decomposition of variance

Mean	0.030
Between	0.110
Within	0.045
Total	0.155

Results

Before formally testing the hypothesis that students with better-performing classmates are more likely to receive special education regardless of their own academic performance, I will present a graphical illustration of the correlation. Figure 2 illustrates the data and the correlation of interest for this analysis in a simplified format. The x-axis is classroom average performance divided into evenly sized categories, and the Y-axis is the percentage of students in special education. The dark bars are the percentage of all students referred and the lighter bars are the percentages of referred students in the lowest-performing quartile of the population. The dark bars follow the pattern that one would expect if there were no inequality in service provision, namely that more students in lower-performing classes are referred to special education. The interesting part is the lighter bars, which show the opposite pattern. A higher share of students performing in the bottom quartile are referred to special education in better-performing classes than in lower-performing classes.

20

Thus, Figure 2 indicates that a low-performing student is more likely to receive special education if she is assigned to a classroom with better-performing students.

Figure 2: Graphical illustration of data
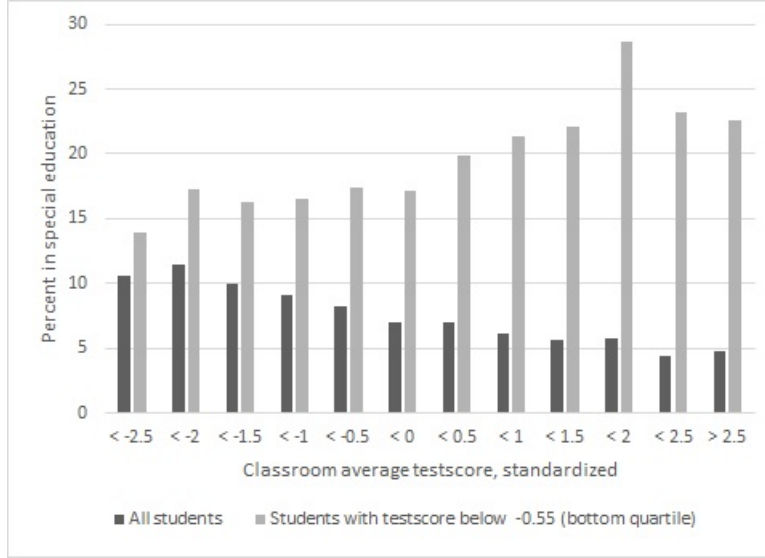
In line with the hypothesis, a formal test of this reveals the same pattern. The results in Table 3 show that a student is more likely to be referred to special education when she has better-performing peers. Since assignment to classes within schools are as good as random, this implies that the likelihood of a student being referred to special education is affected by her classmates, even when the analysis is restricted to variation between classes at the same school. As can be seen in Table 3, Model 1, a student in a class that performs on average one standard deviation better in the national tests is 1.4 percentage points more likely to be referred to special education, given her own academic performance. Considering the baseline that 7.23 % of the students in the population receive special education, this effect size implies that a student in a class that performs one standard deviation worse than the average class is 19.4 % less likely to receive special education than an equally performing student in an average-performing class—a finding that is statistically significant at the 0.001 level. When background information on parental education, gender, immigrant status and the presence

of psychiatric diagnoses at the individual and class level is included in Model 2, it does not change the grouping effect, which further emphasizes that this pattern is not caused by the variation between classes on observable measures.

In Model 3, an interaction term between individual performance and class-level average is included. It is positive, highly significant and about a third of the size of the class average itself. This indicates that the better a student is performing, the more class average matters with regard to the likelihood of special education referral. Or more intuitively, special education referrals for students with lower test scores are less sensitive to the classroom average. This implies that the classroom composition is most important in borderline cases. The needs of very low-performing students will often be met with special education, whereas the referral of less low-performing students will depend to a greater extent on the classroom composition.

As described above, the teacher is the gatekeeper to special education, but whether the student should receive special education, and if so, how much, is based on tests made by a centrally organized PPA unit. The PPA unit covers many schools and many classes. They only meet the individual students and hence they have no knowledge of the student's classmates. Therefore, their assessments can be used to test whether the pattern revealed in Table 3 is simply because more students in better classes need special education, or whether it is a result of an influence of the administrative grouping of students.

Table 4, Model 1 shows that the amount of special education students receive is positively correlated with the average performance of the class, and that this is statistically significant at the 0.001 level. That means, when we look at all students, on average students in better classes receive more minutes of special education. However, in Model 2 the analysis is restricted to students who receive special education, and here the pattern is different. The amount of special education that a special needs student receives does not increase with the performance of the student's classmates; actually it decreases, though the coefficient is not

Table 3: Referrals to special education

	Model 1	Model 2	Model 3
Grouping effect	0.014***	0.014***	0.014***
	(0.001)	(0.001)	(0.001)
Test score	-0.076***	-0.073***	-0.073***
	(0.002)	(0.002)	(0.002)
3rd grade	0.000	0.000	0.000
	.	.	.
4th grade	-0.003	-0.004	-0.004
	(0.002)	(0.002)	(0.002)
5th grade	-0.007***	-0.009***	-0.009***
	(0.002)	(0.002)	(0.002)
6th grade	-0.011***	-0.014***	-0.013***
	(0.002)	(0.002)	(0.002)
7th grade	-0.025***	-0.029***	-0.029***
	(0.003)	(0.003)	(0.003)
8th grade	-0.032***	-0.036***	-0.035***
	(0.003)	(0.003)	(0.003)
9th grade	-0.040***	-0.045***	-0.044***
	(0.003)	(0.003)	(0.003)
Parental education	-0.009***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.001)
Boy		0.013***	0.013***
		(0.001)	(0.001)
Immigrant		-0.018***	-0.018***
		(0.003)	(0.003)
Psychiatric diagnosis		0.088***	0.088***
		(0.004)	(0.004)
Class share, immigrants		0.023*	0.021*
		(0.010)	(0.010)
class share, boys		0.003	0.003
		(0.006)	(0.006)
Class share, psyciatric diagnosis		0.002	0.002
		(0.010)	(0.010)
Class average, parental education		-0.001	-0.000
		(0.001)	(0.001)
grouping effect * testscore			0.004***
			(0.001)
Constant	0.091***	0.088***	0.085***
	(0.002)	(0.011)	(0.011)
N	314504	312590	312590
Schools	1330	1330	1330

Note: School fixed effects, OLS coefficients, robust standard errors in parentheses. *p <
.05; **p < .01; ***p < .001.

statistically significant. Thus, the results in Table 4 imply that students in better-performing classes on average receive more special education, but that this correlation is driven by the over-representation of students from high-performing classes in special education. The special needs students in high-performing classes do not receive more special education than equally low-performing students in other classes. When students with the same academic level do not receive significantly different amounts of special education, this implies that there are not any omitted class-level variables in the analysis that drive the inequality in referrals.

This finding illustrates that when the student's needs are assessed by someone outside the class, the class becomes unimportant; but the classroom seems to matter with regard to whether a student's needs for special education will be assessed in the first place.

There is one caveat to this conclusion that needs some further consideration. Even though the comparison is between classes within the same school, and hence the compared classes draw from the same pool of resources, different classes within the school could be allocated different resources. The Danish law on special education states that special education is given to students whose "development necessitates special consideration or support that cannot be given within the setting of the ordinary class." From the analyses above, it cannot be ruled out that the *ordinary settings of class* are different between lower- and higher-performing classrooms. One could imagine that lower-performing classes are more likely to receive extra resources from the principal to support students' learning within the settings of ordinary class. With more resources to support students inside the classroom, maybe some students who would otherwise need special education could receive the necessary support in class.

As discussed above, there are two possible ways to control for resource allocation between classes with the available data. The first is to include class size as a control variable in the regression, and the second is to limit the analysis to younger students. The academic level of the students is unknown when they start, and because classes typically have the same teacher for at least three years, the teachers in lower grades are expected to be assigned more

Table 4: Amount of special education

	Model 1	Model 2
	Hours of special education pr. student	Hours of special education pr. special education student
Grouping effect	0.018*** (0.002)	-0.007 (0.028)
Test score	-0.103*** (0.005)	-0.121*** (0.026)
3rd grade	0.000 .	0.000 .
4th grade	-0.013 (0.007)	-0.077 (0.061)
5th grade	-0.021*** (0.006)	-0.118* (0.055)
6th grade	-0.041*** (0.006)	-0.244*** (0.063)
7th grade	-0.047*** (0.007)	-0.138 (0.092)
8th grade	-0.069*** (0.007)	-0.316*** (0.083)
9th grade	-0.083*** (0.007)	-0.414*** (0.094)
Constant	0.107*** (0.028)	2.202*** (0.366)
N	307498	17383
Schools	1330	836

Note: School fixed effects, individual and class level variables included, but now shown. OLS coefficients, robust standard errors in parentheses. *$p < .05$; **$p < .01$; ***$p < .001$.

randomly to younger classes. Results of the two analyses are shown below.

In Table 5, Model 1, class size is included in the model. The class size coefficient is negative, but quite small. In line with the argument that teachers may be better able to provide in-class support to special needs students in smaller classes than larger classes, this result implies that students in smaller classes are slightly less likely to be referred to special education. However, the inclusion of class size does not alter the grouping effect: the influence from the classroom composition is the same in smaller and larger classes. In

Table 5, Model 2, the analysis is limited to 3rd grade students. Here, the grouping effect is larger than in the complete sample analysis. Thus, students in 3rd grade are more likely to receive special education in general, and their referral also seems to be influenced by the classroom to a larger extent than older students. These tests are not sufficient to rule out the possibility that lower-performing classes may be compensated in some way that could make referrals to special education unnecessary, but the analyses that can be made to illuminate this explanation do not find any evidence that the referral pattern in Table 3 is not caused by an influence from classroom composition.

Discussion

This study has shown that the administrative grouping of citizens within public agencies may affect public service delivery, such that citizens in better-performing groups end up better off than similar citizens in lower-performing groups. In the special education case studied above, the variation between administrative groups arose from as-good-as-random assignment of students to classes within schools. Despite small differences between classes, this study found a very large effect of administrative grouping. Thus, a student in a class performing one standard deviation lower than an average class was 19.4 % less likely to be referred to special education than an equally performing student in an average-performing class.

In a mapping of the characteristics of students assigned to special education made by Hibel, Farkas, and Morgan (2010), it was also found that students from better-performing schools were more likely to receive special education. This mapping cannot account for selection into schools, and therefore the finding could be a result of unevenly distributed resources. The analyses above are restricted to comparisons of classes within the same school. Resources are allocated at school-level and all classes within the same school draw from the same pool of resources. Hence, differences in available resources to accommodate special education

across classes does not explain the differential thresholds for eligibility across classes. The consistency between the mapping and this study highlights the importance of micro-level explanations.

The results above showed that there was an inequality from the administrative grouping of students when the outcome was partly determined by the teacher, but that this inequality was nonexistent when the outcome measure was determined by someone without knowledge of the class. It is not possible to determine whether this finding is a result of a group-dependent demand, where students or their parents request that the teacher refer a student lagging behind her classmates to special education, or it is a result of a supply-side bias, where teachers' assessments of students' needs are influenced by the class. Either way, this finding raises some more general questions about how much administrative groupings influence service provision, the unfairness of such an influence, and how it can be mitigated.

Administrative groups may often be identifiable on different levels of an organization; in this case the class, the cohort, and the entire school could be considered an administrative group. Given the micro-level arguments presented above, one would expect that group composition has the biggest impact on the lower organizational levels, in this case classrooms. However, street-level bureaucrats may have an awareness of a group on a higher or more abstract level than the classroom group identified here. While teachers teach the same class for a number of years, they often teach several classes at a time, and hence the school level may also influence their perception of how low an academic performance ought to be to necessitate special education. As shown in the decomposition of variance, around 70 % of the variation is between schools. Thus, classes within the same school are more alike than classes in different schools. The school fixed effects design applied controls out all variation between schools, and therefore only investigates a smaller part of the actual variation between groups. Since much of the variance of interest is therefore controlled out, the actual effect from grouping could be substantially larger.

Despite the limitation of the analysis to within-school variation, the influence of the administrative grouping found in this analysis is quite large. One could think of at least two explanations. First, there could be differences in the allocation of resources between classes within schools. This cannot be completely ruled out, but the analysis above indicates that this is not the case. Second, classes where many students are low-performing may adapt to a lower academic level in general, and thereby accommodate the needs of the low-performing students in class without referring them to special education. Despite the reasonableness of this argument, such a procedure may have some distributive consequences that are unwarranted and currently unattended. When students are referred to special education, the agency resources that could have been spent elsewhere are spent to accommodate this student. The analysis above shows that students in better-performing classes are more likely to receive special education, which implies an allocation of resources that does not favor those who need them most.

There may be several ways to mitigate the distributive consequences of grouping within as well as between agencies. First, agencies can be compensated for higher needs among their clients. But, when the unfairness arises at the street level, it is important to note that if equalization transfers are based on local assessments of needs, they will not alleviate the grouping effect, but rather cause an unintended larger compensation to lower-need agencies. Therefore, for equalization transfers to have the intended effect, they need to be made on non-discretionary measures. This taps into a more general discussion of the room for discretion at the front line. When discretion is reduced, the accountability and political control of the agencies increase, but at the cost of front-line workers' ability to maneuver in the complex reality of street-level bureaucracies (Lipsky, [1980] 2010).

Without limiting discretion at the front line, other measures could be taken to minimize the influence of the administrative grouping. If the influence of administrative grouping is caused by the street-level bureaucrats using the group as an anchor, one would expect

28

that the impact of grouping could be drastically reduced if more people were working on the same case, or if the street-level bureaucrat had experience from different administrative groupings. In the case above, an external employee from the PPA unit assessed the needs for special education. This procedure did not alleviate the inequality in referrals, but within the group of students referred to special education, the external assessment of needs was not affected by the administrative grouping of students. Thus, one could assume that a broader experience of different needs and abilities would make street-level bureaucrats less susceptible to the composition of the group. Means to reach a broader frame of reference could be rotation principles or mandatory work placements in organizations with different group compositions. It may also be the case that street-level bureaucrats who are less influenced by the administrative grouping themselves may also be better able to withstand differential demands from citizens that grouping could cause.

Citizens are almost always grouped in some way or another in public service agencies. The presence of a relatively simple and strong pattern like the one showed above, even in an expensive and complex case of service provision, calls for more work into the investigation of whether the administrative grouping of citizens generally influences decisions of public service delivery and to what extent. In addition to the questions raised on administrative groupings, the findings from this study may also call into question our current understanding of differential treatment as something that is rooted in the encounter between front-line workers and citizens.

The assumption in the differential literature has been that the threshold for eligibility is given, but that the evaluation of citizens may be influenced by characteristics of the front-line worker or the citizen, which may cause a different evaluation of similar cases (e.g. Andersen and Guul, 2019; Harrits, 2019; Maynard-Moody and Musheno, 2003; Soss, Fording, and Schram, 2008). The finding from this study suggests that the threshold for eligibility may not be the same across different administrative groupings, and therefore differential treatment

may also occur as a result of the organization itself, independent of individual attributes of citizens and street-level bureaucrats.

If administrative grouping in itself causes differential treatment, a related question emerges. Namely, does an influence from the administrative grouping affect already established conclusions about coping mechanisms and discriminatory behavior at the front line? Harrits (2019) finds that street-level bureaucrats who work in a heterogeneous social context were less likely to apply stereotypes in their evaluation of citizens. This finding implies that there may be some interactive effects between the administrative group and the evaluation of individual citizens. Following Harrits' argument, it could be that the administrative grouping, besides affecting the threshold level for service provision, also affects the stereotypes street-level bureaucrats hold. If this argument is true, citizens with high need for public services may be even more vulnerable to the administrative grouping, because the grouping in itself may increase the threshold for eligibility and at the same time lead to a more stereotypical evaluation of their needs.

Further research is needed to uncover such a relationship between administrative groupings and differential treatment. A better understanding of this phenomenon will advance our understanding of differential treatment in street-level bureaucracies and enable decision makers to ensure that the administrative grouping of citizens is just a tool to provide services and not a cause of inequality.

References

Ammermueller, A., & Pischke, J. (2009). Peer Effects in European Primary Schools: Evidence from the Progress in Iternational Reading Literacy Study. *Journal of Labor Economics, 7*(3), 315–348.

Andersen, S. C., & Guul, T. S. (2019). Reducing Minority Discrimination at the Front Line—Combined Survey and Field Experimental Evidence. *Journal of Public Administration Research and Theory, 29*(3), 429–444.

Andersen, S. C., & Hjortskov, M. (2016). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory, 26*(4), 647–662.

Angrist, J. D., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Beck, N. (2020). Estimating Grouped Data Models with a Binary-Dependent Variable and Fixed Effects via a Logit versus a Linear Probability Model: The Impact of Dropped Units. *Political Analysis, 28*(1), 139–145.

Beuchert, L. V., & Nandrup, A. B. (2014). The Danish National Tests – A Practical Guide. *Institut for Økonomi, Aarhus Universitet. Economics Working Papers No. 2014-25.*

Boadway, R., & Shah, A. (2009). *Fiscal Federalism: Principles and Practice of Multiorder Governance.* Cambridge University Press.

Boone, S., Thys, S., Van Avermaet, P., & Van Houtte, M. (2018). Class composition as a frame of reference for teachers? The influence of class context on teacher recommendations. *British Educational Research Journal, 44*(2), 274–293.

Calsamiglia, C., & Loviglio, A. (2016). Grading on a Curve: When Having Good Peers Is Not Good. *Barcelona GSE Working Paper Series, Working Paper nᵒ 940.*

Davis, J. A. (1966). The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men. *American Journal of Sociology, 72*(1), 17–31.

Einstein, K. L., & Glick, D. M. (2017). Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing. *American Journal of Political Science, 61*(1), 100–116.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? *The Review of Economics and Statistics*, *84*(4), 584–599.

Harrits, G. S. (2019). Stereotypes in Context: How and When Do Street-Level Bureaucrats Use Class Stereotypes? *Public Administration Review*, *79*(1), 93–103.

Harrits, G. S., & Møller, M. Ø. (2014). Prevention at the Front Line: How home nurses, pedagogues, and teachers transform public worry into decisions on special efforts. *Public Management Review*, *16*(4), 447–480.

Helson, H. (1947). Adaptation-Level as Frame of Reference for Prediction of Psychophysical Data. *The American Journal of Psychology*, *60*(1), 1–29.

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who Is Placed into Special Education? *Sociology of Education*, *83*(4), 312–332.

Holm, J. M. (2017). Double standards? How historical and political aspiration levels guide managerial performance information use. *Public Administration*, *95*(4), 1026–1040.

Hoxby, C. M. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working paper 7867*.

Jilke, S., & Tummers, L. (2018). Which Clients are Deserving of Help? A Theoretical Model and Experimental Test. *Journal of Public Administration Research and Theory*, *28*(2), 226–238.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312.

Keiser, L. R. (2010). Understanding Street-Level Bureaucrats' Decision Making: Determining Eligibility in the Social Security Disability Program. *Public Administration Review*, *70*(2), 247–257.

Keslair, F., Maurin, E., & McNally, S. (2012). Every child matters? An evaluation of "Special Educational Needs" programmes in England. *Economics of Education Review, 31*(6), 932–948.

Kvande, M. N., Bjørklund, O., Lydersen, S., Belsky, J., & Wichstrøm, L. (2019). Effects of special education on academic achievement and task motivation: A propensity-score and fixed-effects approach. *European Journal of Special Needs Education, 34*(4), 409–423.

Lipsky, M. ([1980] 2010). *Street-level bureaucracy : Dilemmas of the individual in public services* (30. anniversary expanded ed.). Russell Sage Foundation.

Marsh, H. W. (1984). Self-concept: The Application of a Frame of Reference Model to Explain Paradoxical Results. *Australian Journal of Education, 28*(2), 165–181.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-Fish-Little-Pond-Effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research. *Educational Psychology Review, 20*(3), 319–350.

Maynard-Moody, S., & Musheno, M. C. (2003). *Cops, teachers, counselors: Stories from the front lines of public service.* University of Michigan Press.

Ministry of Finance. (2010). *Specialundervisning i folkeskolen - veje til en bedre organisering og styring* (tech. rep.).

Møen, J., & Tjelta, M. (2010). Grading Standards, Student Ability and Errors in College Admission. *Scandinavian Journal of Educational Research, 54*(3), 221–237.

Neumann, M., Trautwein, U., & Nagy, G. (2011). Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the University entrance qualification in Germany. *Studies in Educational Evaluation, 37*(4), 206–217.

Nielsen, P. A. (2014). Learning from Performance Feedback: Performance Information, Aspiration Levels, and Managerial Priorities. *Public Administration, 92*(1), 142–160.

Olsen, A. L. (2017). Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations. *Journal of Public Administration Research and Theory, 27*(4), 562–580.

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and Moderation - Two Sides of the Same Coin. *American Educational Research Journal, 47*(2), 390–433.

Simon, H. A. (1939). The administrator in search of statistics. *Public Management, 21*, 106–109.

Soss, J., Fording, R. C., & Schram, S. (Eds.). (2011a). *Disciplining the poor: Neoliberal paternalism and the persistent power of race*. University of Chicago Press.

Soss, J., Fording, R. C., & Schram, S. (2011b). The Organization of Discipline: From Performance Management to Perversity and Punishment. *Journal of Public Administration Research and Theory, 21*, i203–i232.

Soss, J., Fording, R. C., & Schram, S. F. (2008). The Color of Devolution: Race, Federalism, and the Politics of Social Control. *American Journal of Political Science, 52*(3), 536–553.

Tummers, L., Bekkers, V., Vink, E., & Musheno, M. (2015). Coping During Public Service Delivery: A Conceptualization and Systematic Review of the Literature. *Journal of Public Administration Research and Theory, 25*(4), 1099–1126.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131.

UVM. (2010). Bekendtgørelse om folkeskolens specialundervisning og anden specialpædagogisk bistand
BEK nr 885.

UVM. (2017). Bekendtgørelse af Lov om Folkeskolen
LBK nr. 989.

Weber, M. (1947). *The theory of social and economic organizations*. Oxford University Press.

Appendix

Figure A1: Distribution of school cohort-specific p-values for Pearson Chi²
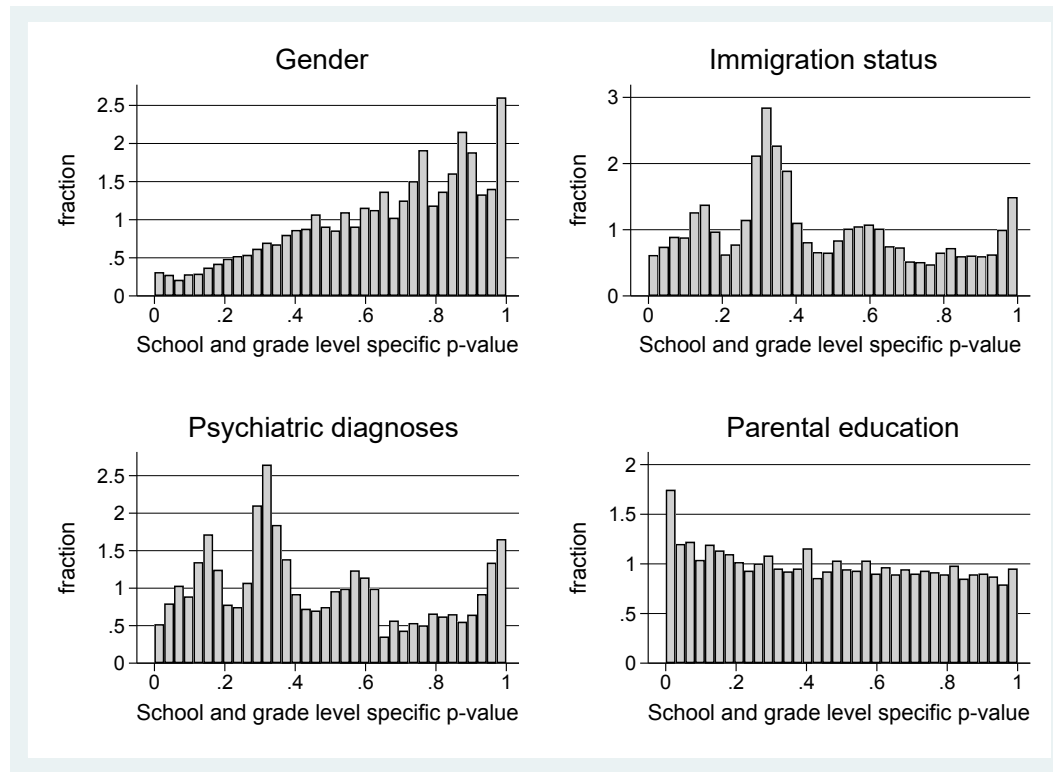
Table 5: Referrals to special education, control for resources

	Model 1 All students	Model 2 3rd grade
Grouping effect	0.014*** (0.001)	0.021*** (0.003)
Test score	-0.073*** (0.002)	-0.086*** (0.003)
Class size	-0.001* (0.000)	
3rd grade	0.000 .	0.000 .
4th grade	-0.004 (0.002)	-0.004 (0.002)
5th grade	-0.009*** (0.002)	-0.009*** (0.002)
6th grade	-0.013*** (0.002)	-0.013*** (0.000)
7th grade	-0.029*** (0.003)	-0.029*** (0.003)
8th grade	-0.036*** (0.003)	-0.036*** (0.003)
9th grade	-0.045*** (0.003)	-0.045*** (0.003)
Constant	0.098*** (0.012)	0.103* (0.042)
N	312590	49653
Schools	1330	1283

Note: School fixed effects, individual and class level variables included, but now shown. OLS coefficients, robust standard errors in parentheses. *$p < .05$; **$p < .01$; ***$p < .001$.

Table A1: Referrals to special education, complete cohorts only

	Referral
Grouping effect	0.019***
	(0.002)
Test score	-0.081***
	(0.003)
3rd grade	0.000
	.
4th grade	-0.001
	(0.005)
5th grade	-0.010*
	(0.005)
6th grade	-0.021**
	(0.007)
7th grade	-0.034***
	(0.006)
8th grade	-0.043***
	(0.008)
9th grade	-0.051***
	(0.007)
Constant	0.136**
	(0.043)
N	66589
Schools	888

Note: School fixed effects, individual and class level variables included, but now shown. OLS coefficients, robust standard errors in parentheses. *p < .05; **p < .01; ***p < .001.

Table A2: Referrals to special education, conditional logistic regression

	Referral
Grouping effect	0.266*** (0.012)
Test score	-1.251*** (0.010)
3rd grade	0.000 .
4th grade	-0.086** (0.028)
5th grade	-0.157*** (0.026)
6th grade	-0.248*** (0.029)
7th grade	-0.534*** (0.029)
8th grade	-0.672*** (0.032)
9th grade	-0.918*** (0.034)
N	254397
Schools	1044

Note: School fixed effects, individual and class level variables included, but now shown. Odds ratios, standard errors in parentheses. *p < .05; **p < .01; ***p < .001.

Table A3: Referrals to special education, Linear probability model estimated with conditional logistic regression sample

	Referral
Grouping effect	0.017***
	(0.001)
Test score	-0.090***
	(0.002)
3rd grade	0.000
	.
4th grade	-0.005
	(0.003)
5th grade	-0.012***
	(0.003)
6th grade	-0.017***
	(0.003)
7th grade	-0.036***
	(0.003)
8th grade	-0.044***
	(0.003)
9th grade	-0.055***
	(0.003)
Constant	0.110***
	(0.013)
N	254397
Schools	1044

Note: School fixed effects, individual and class level variables included, but now shown. OLS coefficients, standard errors in parentheses. *p < .05; **p < .01; ***p < .001.