RIETI Discussion Paper Series 23-E-004

School Accountability and Student Achievement: Neighboring schools matter

MOROZUMI, Atsuyoshi
University of Nottingham

TANAKA, Ryuichi
RIETI

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 23-E-004

January 2023

School Accountability and Student Achievement: Neighboring schools matter[1]

Atsuyoshi Morozumi
University of Nottingham
Ryuichi Tanaka
University of Tokyo / RIETI

Abstract

Previous research on school accountability has shown that the disclosure of school-level results of a national standardized student achievement test has a heterogeneous impact on student achievement across schools. This paper, highlighting a type of standardized test that has no stakes for students (called a national assessment), sheds further light on circumstances under which the disclosure of such information has a desirable impact on student learning. Specifically, utilizing an unanticipated disclosure of the school-level results of Japan's national assessment, which occurred only in one prefecture in 2013, and treating schools in other prefectures as a control group, we show that the information disclosure has a significantly more positive impact on student achievement when the school has a larger number of schools in close proximity (i.e., neighboring schools). The results are robust to the consideration of other possible conditioning factors of the information effect such as school budget autonomy.

Keywords: School accountability, Student achievement, National assessments, Information effects, Neighboring schools

JEL classification: D80, I20, I28.

[1] This study is conducted as a part of the Project "Microeconometric Analysis of Education Policy with Large Administrative Data" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). We are grateful for valuable comments from Yoshiyuki Arata, Daiji Kawaguchi, Keisuke Kawata, Akifumi Ishihara, Masayuki Morikawa, Masaki Nakabayashi, Daigo Nakata, Masahiro Shoji, Richard Upward, Eric Weese, and participants at the project meeting and seminar at RIETI, Kansai Labor Workshop, and seminar at the University of Tokyo. We are also indebted to Yumiko Mitarai for her assistance in data collection. We acknowledge the financial support for this project from the Japan Society for the Promotion of Science (Grant Numbers 20H05629 for Ryuichi Tanaka).

1 Introduction

Many countries conduct a national standardized test of education performance. While the exact purposes of conducting such tests vary across countries, OECD (2015) categorizes them broadly into two types: national assessments and examinations. The former is a type of standardized test conducted mainly to provide schools with student diagnostic information and offer formative feedback to parents, without having any stake for students. The latter type is instead used to certify the current education level or to select students to a higher education level, thus having a high stake for students. One key design feature of such standardized tests, irrespective of the types, is the level of aggregation at which the results are disclosed to the public; it could be the school level or a higher level such as municipality, province, or state. For example, in 2015, 31 (out of 38) OECD countries conducted a national assessment at the primary education level, of which 14 published the results at the school level and the remaining at the higher level of aggregation (OECD (2015)).[1]

The growing strand of the literature on school accountability has examined the relevance of the key feature of a national standardized test to student outcomes, highlighting the effects of the disclosure of the *school-level* results both in the context of national assessments and examinations (e.g., Koning and van der Wiel (2012), Burgess et al. (2013), Camargo et al. (2018), Morozumi and Tanaka (2020), and Cilliers et al. (2021)). The general consensus is that the information disclosure itself, even when the school's results are not linked to explicit consequences such as monetary awards/penalties and takeover threats, has a substantial potential to improve student achievements.[2] At the same time, the related works have indicated that the disclosure effects are *heterogeneous* across schools. For example, Koning and van der Wiel (2012) and Burgess et al. (2013) suggest that the effects are greater when schools' initial performance was relatively poor, since such schools respond to accountability pressure better. Camargo et al. (2018) show that the effects are larger when schools are private, since school managers in such schools are subject to market incentives.

[1] Primary education typically begins at the ages five, six or seven and lasts for four to six years.

[2] A separate strand of the school accountability literature examines the effect of accountability system based on such explicit consequences (including Hanushek and Raymond (2005), Chiang (2009), Rockoff and Turner (2010), and Rouse et al. (2013)). See Figlio and Loeb (2011) for a review.

This paper sheds further light on conditions under which the information disclosure of school-level results of a national assessment has a positive impact on student achievements. Specifically, we test the conjecture that the more *neighboring schools* there are (i.e., the more schools exist in close proximity), the more the information disclosure improves student achievement. We motivate this conjecture by presenting a simple principal-agent model where the principal is parents and the agent is a school, and the school's effort level to produce output is its private information.[3] The model abstracts away the student's incentive, since a national assessment, by definition, does not have stakes for students. The disclosure of school-level results of the standardized test helps parents monitor the school's effort level by providing a proxy of the school's output that is *comparable* across schools.[4] Here, to the extent that it is external comparison with neighboring schools (parents typically know about) that facilitates parents' monitoring of the school's effort, the presence of neighboring schools mitigates the moral hazard problem better, eliciting more effort from the school. The model thus suggests that when there are more neighboring schools, the disclosure of school-level results of national assessments improves student achievements *even further*.

One of the well-known challenges to estimate a causal effect of school accountability is lack of an adequate control group for the counterfactual (e.g., Figlio and Ladd (2007)). To address this challenge, we utilize an unanticipated disclosure of school-level results of a Japan's national assessment occurred only in one of the 47 prefectures in 2013 at the primary education level, while regarding schools in the remaining prefectures as a control group. This disclosure was a purely unanticipated and exogenous event for the following two reasons. First, the disclosure of school-level results in the national assessment was simply *not* allowed that time throughout Japan by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the central education authority in Japan. Second, this disclosure was enforced not by the local education authority (called an education board), but by the then governor of the prefecture, who was *not* supposed to interfere

[3] The model is based on Ferreyra and Liang (2012).

[4] Bergbauer et al. (2021) also refer to the principal-agent framework to motivate the effects of a standardized test that allows for external comparison of school performances. Using student-level data from 59 countries, they show that the use of such a standardized test within a country promotes student performance in Programme for International Student Assessment (PISA) test in the initially low- and medium-achieving countries.

directly with the prefectural education policy. This utterly unprecedented event, occurred outside of the then regulatory framework, provides a natural experiment to estimate the causal effects on student outcomes of school accountability via information disclosure, highlighting neighboring schools as a key conditional factor. We call this exogenous event an *accountability shock* below.

We find that the disclosure of information on school-level results of the national assessment generally promotes student achievements at the primary education level. More importantly, however, we also find that the effects on student achievements are significantly greater when there are more schools in close proximity, i.e., neighboring schools. Quantitatively, when the school has one more neighboring school within the 1.5km radius, the disclosure of school-level results of Japanese (a subject whose school-level results were published in the prefecture) improved the Japanese score of an individual student in the subsequent assessment by about 0.04 standard deviations. The results on the relevance of neighboring schools to the information disclosure effect remain robust even after other possible conditioning factors of the effect such as school budget autonomy are taken into account. Moreover, the greater number of neighboring schools creates the positive information effect on student achievements only when schools are disclosed as low-performing schools. This result suggests that parents do not attempt to hold accountable schools that are disclosed as a high-performing schools to begin with.

This paper contributes to the growing strand of the school accountability literature that highlights information effects of school-level results in a national standardized test. For example, Burgess et al. (2013) exploit a natural experiment setup where, in both Wales and England, school-level results of national examination at the secondary education level had been published until 2001, while only Wales *abolished* the disclosure in that year. They show that this shutdown of information disclosure in Wales worsened the examination performance of the schools relative to English schools, claiming that information disclosure of school-level results of national examination has the potential to increase school effectiveness. Camargo et al. (2018) study how, in Brazil, the information disclosure of school-level results in the national examination at the end of secondary education affected the school performance in the test. They find that the information disclosure had a greater positive impact on scores of private schools than on scores of public schools. Koning and van der Wiel (2012) show that, in the Netherlands, the publication of quality ratings of secondary schools that are partly based on the school's performance in a centralized examination induced a positive reaction from schools to improve their performance.[5] Unlike the aforementioned papers that examine the information effect in the context of national examinations that have high stakes for students by definition, Morozumi and Tanaka (2020) consider such an effect instead in a national assessment that provides schools with student diagnostic information without having stakes for students. They show that, in Japan, publishing the school-level results of a national assessment increases students' test scores across the entire score distribution, with no evidence of adverse impacts on non-cognitive skills, and further find that the information disclosure prompts schools to utilize diagnostic information from the assessment to improve teaching quality.

While the information disclosure of school-level results of a national standardized test generally promotes student achievement, the effects are known to be heterogeneous across schools. As mentioned, previous works indicate that effects are greater for schools whose initial performances were rather poor, because they tend to respond to the information disclosure more strongly in light of the possible loss of reputation (e.g., Koning and van der Wiel (2012), Burgess et al. (2013), Morozumi and Tanaka (2020), and Cilliers et al. (2021)). Camargo et al. (2018) also show that effects are greater for private schools where school managers are subject to market incentives. Further, albeit not in the context of accountability effects based on the disclosure of school-level results of a standardized test, works such as OECD (2011) and Hanushek et al. (2013) demonstrate that school accountability exhibits a significant interaction effect with school autonomy. For example, Hanushek et al. (2013) show that school autonomy has a more positive impact on student achievement when centralized exit examinations are in place as an accountability device.[6] Motivated by such results, our analysis of the role of neighboring schools in the information disclosure (i.e.,

[5]In the context of developing countries, Cilliers et al. (2021) show that, in Tanzania, public rankings of school performance in the national examination at the primary education level prompted poorly-performed schools to improve the performance by excluding students strategically from the testing pool.

[6]Central exit examinations can be regarded as an accountability device because they signal student achievement to potential employers on the labor market and also to the higher education stage, thereby increasing students' reward for learning and parents' incentive of monitoring the education process (e.g.,Bishop (2006)).

accountability) effect controls for an interaction between the disclosure and school budgetary autonomy.[7] We further control for interaction effects between the disclosure and other factors such as parents' income levels and the school-principal's tenure.[8]

The rest of the paper is organized as follows. Section 2 utilizes a principal-agent model to motivate the main testable conjecture on the relevance of neighboring schools to the information disclosure effects of a national assessment on student achievements. Section 3 provides the background of the analysis, detailing the said unanticipated information disclosure occurred in one prefecture in Japan. Section 4 explains an empirical methodology, and Sections 5 describes the data. Sections 6 and 7 present and discuss results. Last, Section 8 offers concluding remarks.

2 Theoretical motivation

This section motivates the empirical analysis below, by deriving a testable conjecture on the relevance of neighboring schools to the information effect of disclosing school-level results of a national assessment. Specifically, we do this using a simple principal-agent model where the principal is parents and the agent is a school, and the school's effort level to produce student achievement is its private information.

2.1 Setup

The model is based on Ferreyra and Liang (2012). The key element common to their model is the reduced-form way in which a moral hazard problem between parents and schools is introduced (as elaborated below). Our innovation here is to consider the relevance of the number of neighboring schools to the effects of disclosing school-level results of a national assessment. Different from Ferreyra and Liang (2012) who differentiate the incentive structure of public from private schools, we only consider public primary schools. This is in line with the reality that in Japan there are very few private schools at the primary level (see Section 3). Although the role of the government is specified, the action is entirely exogenous.

[7] Budgetary autonomy is one of the three key school autonomy components examined by Hanushek et al. (2013). The other two are academic-content autonomy and personnel autonomy.

[8] Burgess et al. (2013) also consider an interaction between the information disclosure and parents' income levels.

The model is a one-period model. The sequence of events within the period is as follows:

1: The government exogenously decides two policy parameters for (public) schools: the school's effort standard to produce student achievement along with the level of funding.

2: Parents, having rational expectations, choose the level of effort to monitor the school.

3: Schools, conditional on the parents' monitoring level, decide their actual effort level.

The model abstracts away the student's incentive and only considers the interaction between the school and parents, since our focus is on a national assessment that, by definition, does not have stakes for students.[9]

2.1.1 Parents' behaviour

Once the government has moved first, parents choose the monitoring level to maximize their utility. The utility function takes the form of:

$$U_{s,p} = \alpha y_{s,p} - \{1 - q[\rho_s(n_s)]\} m_{s,p}, \tag{1}$$

where

$$y_{s,p} = F(e_s) = a_s e_s + \varepsilon_s. \tag{2}$$

In Eq.1, $U_{s,p}$ represents the utility of parents, p, whose child attends school s. In the right hand side, $y_{s,p}$ is the student (child)'s achievement, which increases parents' utility. $m_{s,p}$ represents the parents' effort exerted to monitor the school, which carries a utility cost. The key assumption here is that the marginal cost of monitoring effort, $1 - q[\rho_s(n_s)]$, depends on the following two factors.[10]

[9] More broadly, the structure of education system is characterized by complex strategic interactions among different players that include not only schools, parents, and their children, but also school administrators at different levels (see, e.g., Bergbauer et al. (2021) for general discussion). In the Japanese context, school administrators include the local (municipality-level) education authority, called a municipal education board (explained below in Section 3), and there are likely to be parent-education authority as well as school-education authority information and monitoring problems too. Thus, while the present theory (and paper) focuses on how disclosure of school-level results in a national assessment may affect student achievements by helping shape the parents-school information problem, such disclosure can possibly affect them through other interactions involving the education authority too.

[10] Ferreyra and Liang (2012), modeling heterogeneous households in terms of income and ability, assume that the marginal cost of effort is higher when household ability is lower.

One is whether school-level results of a national standardized assessment are disclosed to the public. Specifically, q, a dummy variable, takes the value of one if school-level results of such an assessment are published, and zero if they are not. Another, related, factor is the number of neighboring schools, n_s. When school-level results are disclosed (i.e., $q = 1$), an increase in the number of neighboring schools reduces the marginal cost of monitoring, as denoted by $\rho_s(n_s)$ being an increasing function of n_s, with $0 < \rho_s(n_s) < 1$. Put simply, while the disclosure of standardized school-level results facilitates parents' monitoring of the school via external comparison of results across schools, the presence of more neighboring schools reduces the marginal cost of monitoring *even further*. This assumption is justifiable because parents are typically familiar with characteristics of neighboring schools through interactions with parents from those schools, and thus are in a better position to monitor the school via external comparison with them.[11] If, in contrast, parents are only able to compare results with distant schools they know much less about, such comparison is unlikely to provide them with much information about the school's effort.[12]

Eq.2 is a simple production function of student achievement, $y_{s,p}$. The function is increasing in the product of a_s and e_s, where e_s is the school's effort level, and a_s is the composite of all the predetermined factors relevant for the school that affect the achievement (such as the family environment, school resources such as class size, and public education spending per student). Last, ε_s is an idiosyncratic (i.e., school-specific) shock with zero mean.

[11]For example, students in Japan typically attend tutoring schools in the evenings and/or weekends (in our sample below, 51% of primary-school students use tutoring schools). There, parents of the students from different schools within the region interact and learn about what is happening in neighboring schools.

[12]Our interest in the role of the number of neighboring schools is motivated purely as a mitigating factor of information and monitoring problems between parents and schools. However, in the literature of school accountability, the number of neighboring schools has been used as a measure of local competition (e.g., Burgess et al. (2013)). Acknowledging this, it is worth noting that a system of school choice is not common in Japan particularly at the primary education level. In particular, in the prefecture where the accountability shock hit in 2013, school choice was essentially absent, suggesting that parents could only send their child to the school within the district of their residence. Therefore, together with the fact that the overwhelming majority of primary schools in the prefecture were public and thus not subject to market incentives, the lack of school choice implies that local competition among schools was likely to be weak (even if present) regardless of the number of neighboring schools.

2.1.2 Schools' behaviour

Schools decide their effort level, e_s, such that their utility is maximized. Their decision is conditional on parents' monitoring effort, $m_{s,p}$, determined before their move. Schools' utility, U_s, is expressed as:

$$U_s = \chi_s - \mu e_s - \frac{\beta m_{s,p}}{2}(\overline{e} - e_s)^2, \tag{3}$$

where exogenously-determined government's funding, χ_s, increases utility, while effort, e_s, is decreases it. The last component of the right hand side captures moral hazard in a principal-agent model in the manner of Ferreyra and Liang (2012). Specifically, with $\overline{e}$ being the standard effort level set by the government, the component describes the situation where as parents conduct more monitoring, under-provision of the school effort is punished more, e.g., through the loss of reputation. This way, we formalize that parents' (principal's) monitoring disciplines the school's (agent's) incentive to under-provide effort. To note, the case where schools over-providing effort ($e_s > \overline{e}$) is ruled out simply because it is not consistent with a typical moral hazard problem.

2.2 Solution

Solving backwards, we first consider the school's choice of actual effort conditional on parents' monitoring level, $m_{s,p}$. Differentiating Eq.3 with respect to e_s and setting it to zero yields:

$$e_s = \overline{e} - \frac{\mu}{\beta m_{s,p}}. \tag{4}$$

This equation says that when parents' monitoring level is higher, the actual effort, e_s, is closer to the expected effort, $\overline{e}$, indicating that the principal's monitoring disciplines the under-provision of the effort from the agent.

Turning then to the parents' decision on how much effort they exert to monitor the school, we assume that parents are rational in the sense that they utilize all the information available. Thus, substituting Eq.2 into Eq.1, incorporating that the expected value of idiosyncratic shock is zero

($E(\varepsilon_s) = 0$), and combining with Eq.4, we have the expected utility of parents as:

$$E(U_{s,p}) = \alpha a_s \left(\bar{e} - \frac{\mu}{\beta m_{s,p}} \right) - \{1 - q[\rho_s(n_s)]\} m_{s,p}. \tag{5}$$

Assuming further that parents are risk neutral, they choose their effort level, $m_{s,p}$, to maximize the expected utility. The equilibrium monitoring from parents is thus obtained as:

$$m_{s,p} = \left\{ a_s \frac{\alpha\mu}{\beta\{1 - q[\rho_s(n_s)]\}} \right\}^{1/2}. \tag{6}$$

Last, combining Eq.6 with Eq.4 gives the equilibrium effort of the school, which, in turn, can be substituted into the production function (Eq.2) to yield the equilibrium student achievement as:

$$y_{s,p} = a_s \left\{ \bar{e} - \frac{\mu}{\beta} \left\{ a_s \frac{\alpha\mu}{\beta\{1 - q[\rho_s(n_s)]\}} \right\}^{-1/2} \right\} + \varepsilon_s. \tag{7}$$

Accordingly, the main messages from Eq.7 are the following. When information on school-level results of a national assessment are available and thus enable external comparison, i.e., $q = 1$,

$$\frac{\partial y_{s,p}}{\partial n_s} > 0. \tag{8}$$

In the meantime, when such information is not publicly available, i.e., $q = 0$,

$$\frac{\partial y_{s,p}}{\partial n_s} = 0. \tag{9}$$

These equations, combined, say that when the school-level results of a national assessment are disclosed to the public, an increase in the number of neighboring schools (i.e., schools in close vicinity of school, s) leads to a rise in the student achievement. This is the conjecture we test empirically below. To recap, intuition is the following. It is the presence of more neighboring schools that particularly facilitates parents' monitoring of the school via external comparison of school-level results. When monitoring is less costly, parents increase monitoring of the school such that under-provision of its effort is disciplined. The increase in the school effort, in turn, increases the student achievement for a given level of inputs and a realized idiosyncratic shock.[13]

3 Background

This section first introduces the national assessment in Japan that has been implemented annually since 2007. Then, after providing some background information on the prevailing disclosure system of the national assessment results as of 2013, we explain an utterly unexpected and exogenous disclosure of school-level results of the assessment that occurred in one of the 47 prefectures in 2013. This exogenous event provides a natural experiment to estimate the causal effects on student outcomes of school accountability via information disclosure, highlighting the role of neighboring schools as motivated by the theory above.

3.1 National Assessment of Academic Ability

Japan's compulsory education consists of 6 years of primary education and 3 years of lower secondary education, spanning the ages 6 to 15. Since 2007, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has conducted an annual national standardized assessment called National Assessment of Academic Ability (NAAA). The NAAA takes place typically in late April, the month when school year starts in Japan. Students in the final years of public primary and lower-secondary schools, corresponding to 12 and 15 years old, sit the assessment. Practically all (99.9% to 100%) of such schools throughout Japan participate.[14,15] The NAAA thus covers the majority of students at the relevant ages: students in public primary and lower-secondary schools, excluding their private counterparts, accounted for roughly 98% and 92% of the total number of students in 2019, respectively. The subjects assessed have changed somewhat year by year, but Japanese language (Japanese hereafter) and Mathematics have been assessed every year, both

[13]Regarding the noise term of ε_s, this does not have any impact on the key results of Eqs. 8 and 9. We include the term just to avoid the situation where parents can deduce the actual level of school effort without any error, despite the assumption that it is the school's private information.

[14]In the years when the participation rate was 99.9% (instead of 100%), it was because a locally-happened event such as a natural disaster and spread of some infectious disease prevented the affected schools from taking part.

[15]As an exception, in 2010 and 2012, the NAAA only covered randomly sampled students (about 30% of the total), and in 2011 and 2020, it was cancelled altogether due to the Great East Japan Earthquake and the COVID-19 pandemic.

at the primary and lower-secondary levels. Both subjects were further divided into "Basics" and "Applications" components until the 2018 NAAA.[16]

The aim of the NAAA is to improve students' learning environment by providing *diagnostic feedback* to schools and parents of the students. It is indeed a national assessment defined above in Section 1, since it does not have any bearing on students' progression through school or certification. Besides, schools' performance in the NAAA does not have any explicit consequences such as funding loss and takeover threats for schools, *unlike* a test-based accountability system created, for example, by the No Child Left Behind Act of 2001 in the US where the fraction of students in a school who achieved proficiency in a particular subject was linked to a set of sanctions. In the NAAA, any consequences for schools are thus implicit, likely associated with their reputation. This is the basis on which parents' monitoring effort enters directly into the objective function of the school in Eq.3 above, rationalizing why parents' monitoring incentivizes the school to exert more effort to improve its performance in the NAAA such that it avoids possible reputation loss.

3.2 Accountability shock and natural experiment

This paper utilizes an unanticipated disclosure of school-level results of the NAAA in one prefecture of Japan in September 2013 for causal inference. To describe the background of this natural experiment, we first explain how results of the NAAA were *supposed* to be disclosed back then.

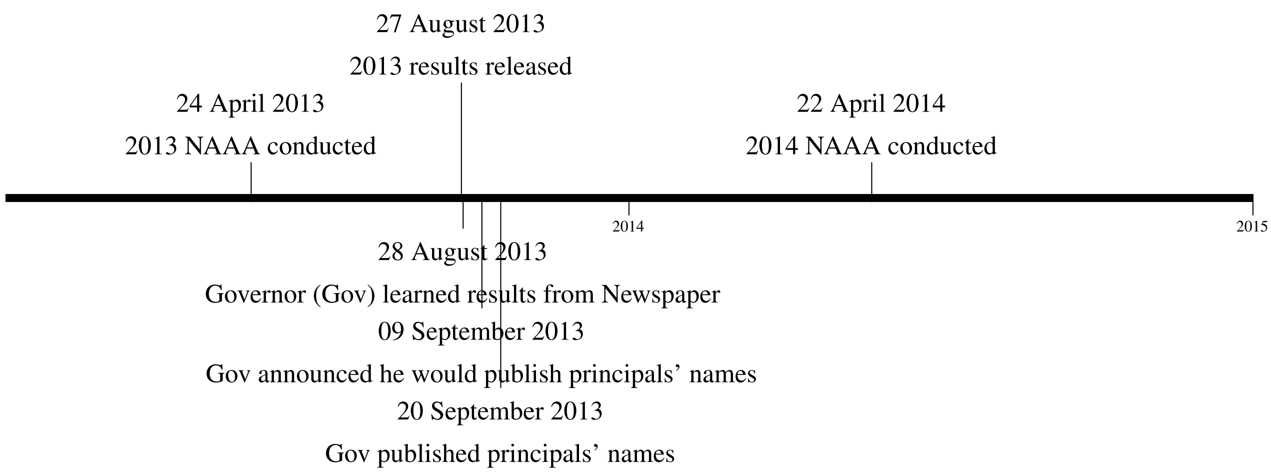
There are (and were) 47 prefectures and about 1,700 municipalities that are subordinate to the prefectures.[17] In each of the prefectures and municipalities, an education board, an executive body of usually 5 members, designs and implements education policies locally; governors of prefectures and mayors of the municipalities are *not* involved in the process directly.[18] For example, a prefectural education board appoints and allocates teachers to all the public schools within the prefecture, whereas a municipal board establishes/abolishes public primary and lower secondary schools within the municipality, and manages key aspects of school operations such as textbook

[16] "Basics" primarily tests the acquisition of basic knowledge of the subject, whereas "Applications" tests the student's ability to apply the basic knowledge to wide range of real-world problems.

[17] The exact number of municipalities varies slightly year by year when some merge.

[18] To be exact, governors and mayors still had some indirect influence on education policies via the budget allocation of public spending. The bottom line, however, is that education boards retain discretion in education policymaking.

Figure 1: Accountability shock: Sequence of events in prefecture A over the 2013-14 period



Note: New school year starts on 1 April in Japan.

selections. Turning to the specifics of how NAAA results were disclosed back in 2013, by regulation, a prefectural education board only had discretion to publish the prefecture-level results, and *not* the municipality- and school-level results. Meanwhile, a municipal board was allowed to publish the municipality-level result, but *not* the school-level results. The upshot here is that together with the fact that governors and mayors were not supposed to be involved in education policy-making directly, there was *no* expectation whatsoever that school-level results of the 2013 NAAA implemented in the April would be disclosed to the public.

However, after the 2013 NAAA results were released by the MEXT to municipalities throughout Japan on 27 August that year, the then governor of one prefecture (called prefecture A hereafter) decided to disclose information on school-level results of the assessment. The timeline of events is summarized in Figure 1. On the next day, 28 August, the governor learned results from a newspaper (not from the prefectural education board) that in one of the subjects covered, specifically Japanese (Basics), the prefecture mean score of the 6th grade students (who were in their final year of the primary education level) was lowest among the 47 prefectures. In response, they made a hugely controversial announcement on 9 September that they would disclose the names of principals whose schools scored below the national average in that subject, to make them accountable for their schools' performances. Then, on 20 September, they actually published on the website of the governor's office the names of the principals whose school scored *above* the national average in that subject. Since school principals' names were (still are) public information in Japan, the consequence, irrespective of the choice of disclosing the names below or above the national average, was that primary schools in the prefecture were categorized publicly according to the school-level performance in the particular subject relative to the national average.

This is where this unanticipated event provides a natural experiment to estimate a causal inference of the accountability effect via the information disclosure on student outcomes. The event, we call the accountability shock, happened utterly outside the education system such that no one expected at the time of the 2013 NAAA in April that this would come. Indeed, the disclosure was enforced purely based on the governor's sole judgment without the prefectural education board's and the MEXT's approval, nor with the agreement of school principals. Here, acknowledging that after the disclosure on 20 September 2013, still 7 months were left till the next NAAA on 22 April 2014, we estimate the effect of this accountability shock on students' test scores in the 2014 NAAA, while paying particularly attention to the relevance of neighboring schools to the effect as the aforementioned theory suggests (Section 2). Last, since this shock occurred only in Prefecture A, we regard schools in the remaining prefecture as a control group that is necessary to identify the causal effect of the information disclosure.

4 Empirical methods

The empirical analysis essentially follows a differences-in-differences (DD) approach, utilizing the fact that the accountability shock created the situation in which school-level results of schools in Prefecture A alone were disclosed to the public. We use schools in the other prefectures as a control group. Our focus below is on the short-run effect of the shock, investigating how the shock in September 2013 affected test scores of the next NAAA in April 2014.[19]

[19]We focus on the short-run effect because only the 2013 information disclosure was a pure unanticipated event. To clarify, after results of the 2014 NAAA were released in the late August, the same information (i.e., the names of primary school principals whose schools averaged above the national average in Japanese, Basics) ended up being published by the then (same) governor of Prefecture A. Although this disclosure was again controversial and still out-

4.1 Empirical model

To estimate the effect of the accountability shock on test scores and the relevance of neighboring schools to the effect, we use the following regressions equation:

$$\begin{aligned} Y_{ism,t} = {} & \beta_1 + \beta_2 \, Treat_t + \beta_3 \, Treat_t * N_{sm} + \beta_4 \, Treat_t * \overline{Y}_{sm,t-1} + \beta_5 \, \overline{Y}_{sm,t-1} \\ & + X_{ism,t} \, \eta + Z_{sm,t} \, \theta + C_{m,t} \, \kappa + \upsilon_s + \varepsilon_{ism,t} \end{aligned} \tag{10}$$

where $Y_{ism,t}$ represents the standardized test score at the national level of 6th-grade student i in public primary school s located in city (municipality) m of Prefecture A in the NAAA implemented in year $t(= 2013, 14)$. Specifically, $Y_{ism,t}$ is either the student's score of Japanese (Basics) or Total score (i.e., the average of the four components: Basics and Applications of Japanese and Mathematics). As noted, the former is the particular component whose school-level results of the 2013 NAAA were disclosed to the public in September 2013. $Treat_t$ is a time dummy (and thus treatment dummy in our context) taking the value of one in year 2014. N_{sm}, which is time-invariant, is the number of neighboring schools within 1.5 km radius. We consider different radius sizes for robustness.[20] The interaction term between $Treat_t$ and N_{sm} is included to estimate the effect of the information disclosure conditional on the number of neighboring schools. $\overline{Y}_{sm,t-1}$ is the lagged average test score at the school level, to address the possible mean-revision, whereby a school with unusually low initial score might see larger improvements than the other schools regardless of the treatment.[21] The interaction between the treatment dummy and the lagged school-mean score is motivated by the possibility that the information disclosure (accountability) effect might be stronger when the initial score is lower. As noted, robustness checks further consider other possible conditional factors of the information effect at the school level, such as the school's budget autonomy and parents' income level.

side the regulation, the disclosure was likely to be expected at least partially. To add, after 2015, the same information has never been disclosed for the prefecture.

[20] The use of 1.5 km as a reference is motivated by the fact that the mean (median) of the distance to the nearest school in our sample is 1.63 (1.83) km.

[21] Since the 2012 NAAA was sample-based (covered about 30% of eligible students throughout Japan), for 2013, we use the latest possible school-means available as the lagged school-mean scores, which could be either 2010 or 2009 scores (in 2011, the NAAA was canceled, and in 2010 it was again sample-based; see footnote 15 for details).

$X_{ism,t}$ is a vector of variables that reflect the student's characteristic and home environment, covering their sex, whether they eat breakfast regularly, wake up at a regular time, sleep at a regular time, and attend a tutoring school (private supplementary school).[22] $Z_{sm,t}$ contains variables related to school environment, specifically, class size in the previous year (when student i was in the 5th-grade), average years of teaching experience of teachers, and the proportion of students who receive financial support, which is inversely related to the income level of the parents.[23] $C_{m,t}$ contains expenditure and institutional variables at the municipality level. Education expenditure is converted to per student, averaged over the past 3 years (corresponding to the time when students were between the third to fifth grade), and log-transformed. Institutional features of education systems are the frequency of school inspections (by municipal education board members) and the degrees of school autonomy in terms of budget allocation. To note, parental school choice is absent in our sample, meaning that parents can only send their child to the school within the district of their residence. υ_s is school fixed effects, capturing time-invariant, unobserved school-specific factors.

4.2 Identification strategy

Our identification strategy builds on the DD estimator with Eq.10, exploiting the fact that the accountability shock created the situation in which school-level results were disclosed in all municipalities in Prefecture A, but not in the other prefectures. We consider schools in Prefecture A as treated schools, and use schools in the other prefectures as a control group. Specifically, note that our dependent variables are test scores standardized at the national level. With the fact that the share of students in schools in Prefecture A consists of only about 3 % of 6-graders in public primary schools throughout Japan, the standardized test score at the national level is effectively the difference between the score of students in treated schools and the average score of students across the remaining prefectures as control.

[22] We regard the student's attendance to a tutoring school as a proxy for the household's socio-economic status. With high socio-economic status, students tend to be sent to a tutoring school for extra education.

[23] Municipalities provide financial support for students from low-income households. Hence, the proportion of students with school financial support is inversely associated with the income level of parents of the school district.

5 Data

Table 1 reports descriptive statistics of the sample of cities in Prefecture A, which is located west of Tokyo and populated by about 3.64 million people. The dataset covers the total number of 60,917 6th-grade students who attended pubic primary schools for the 2013-2014 period. We have repeated observations at the school level rather than the student level. All test scores from the NAAA are standardized within subject and year at the national level with the mean of 50 and the standard deviation of 10. The averages of Japanese, Basics, and Total score (covering Japanese and Mathematics, both Basics and Applications) across students and years are both lower than 50, meaning that the average scores of Prefecture A is lower than the respective national averages. This partly reflects the fact that the prefecture average score of Japanese, Basics, in particular, was the lowest among all the prefectures in 2013.

Regarding the student's characteristic and environment, when sex takes the value of 1 (2), that means that they are female (male). The proportion of the female students is slightly lower than 0.5. Next, three of the student's home environment variables (on breakfast, waking up, and sleeping) are ordinal variables, where the value of 3 (maximum) means that their answer to the question is "yes, certainly", and the value of 0 (minimum) means "no, not at all". Thus, the average of 2.86 for the question asking whether they eat breakfast every morning indicates that a majority of students eat breakfast every morning. Averaged responses to the questions on whether they wake up/go to bed at about the same time every day are relatively low along with the larger standard deviations, that indicate that there are large variations in their responses. The use of a tutoring school is a dummy variable, which takes the value of one if they use it, and zero otherwise. The mean of 0.51 indicates that 51 percent of students use a tutoring school service outside of school.

Turning to school environment variables, on average, there are 1.22 (neighboring) schools within the 1.5 km radius. The standard deviation is large (1.3), with the maximum (minimum) of 7 (0). The mean of the total number of students per primary school is 580. 5.91 percent of students receive school financial support, though there is substantial heterogeneity across schools with the maximum (minimum) of 25 (0) percent. The average class size of the fifth grade in the

Table 1: Descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.	Nat. Mean
Standardized test scores[a]					
Japanese (Basics)	48.72	9.93	14.39	68.09	49.95
Total: Avg of Japanese and Math (Basics, Apps)	49.67	8.5	19.14	66.79	49.93
Student's characteristic and environment					
Sex[b]	1.49	0.5	1	2	1.49
Eat breakfast everyday[c]	2.86	0.45	0	3	2.84
Wake up at a regular time[c]	2.45	0.71	0	3	2.48
Sleep at a regular time[c]	2.11	0.8	0	3	2.13
Use a tutoring school[d]	0.51	0.5	0	1	0.49
School environment					
Number of schools within 1.5km radius	1.22	1.3	0	7	n.a.
Total number of students per school (100 students)	5.8	2.43	0.31	12	5.05
Prop of students receiving financial support[e]	5.91	4.41	0	25	13.78
Class size in previous school year	30.42	4.33	10	47.67	30.68
Prop of teachers with experience under 5 yrs[e]	18.78	7.71	0	50	21.82
Prop of students needing language help[e]	0.57	1.5	0	14.29	0.27
School principal's tenure[f]	0.65	0.48	0	1	n.a.
City-level education expenditures and institutions					
Log of educ exp per primary student (1,000 yen)[g]	4.77	0.27	4.34	5.92	n.a.
School inspection[h]	0.73	1.38	0.06	10.67	n.a.
Budget autonomy[i]	0.25	0.44	0	1	n.a.
Number of observations		60,917			2,171,245

Notes: Statistics over the 2013-14 period, covering 23 cities in Prefecture A. (a) Test scores are standardized within year at the national level with the mean of 50 and the standard deviation of 10. Total score is the average of Japanese and Math (both of which are further divided into Basics and Applications). The Basics (Applications) sub-component tests the acquisition of basic knowledge of the subject (the ability to apply the basics to real-world problems). (b) 1 (2) means female (male). (c) 3 (max) means yes, certainly; 0 (min) means no, not at all. (d) 1 (0) means they attend (do not attend) a tutoring school. (e) Measured in percent. (f) 1 (0) means that the principal has more (less) than 1 year of experience at the school. (g) Average expenditure over the previous 3 school years. (h) Number of school visits by board member(s) per school in the previous school year. (i) If schools were given autonomy in budget allocations in a given school year, the value of 1 is given, 0 otherwise. Numbers are based on the average over the previous three school years (in line with expenditure variables).

previous school year is 30.4. The use of the previous-year figure is motivated by the fact that the NAAA takes place in the first month (i.e., April) of the school year. The proportion of teachers with teaching experience of fewer than 5 years is 18.8 percent on average. 0.57% of students need help in terms of language (Japanese), although, as suggested by the standard deviation (1.5%), many schools have a greater proportion of such students. School principal tenure takes the value of one when the principal has an experience of more than one year at the school and zero otherwise. The mean of 0.65 means that about one-third of the principal is relatively new with a less-than-a-year experience at the school.

We lag expenditure and institutional variables considering that the NAAA takes place in the first month of the school year. The expenditure data is divided by the total number of primary school students in the city (including 1st to 6th grades). To capture possibly delayed effects on student outcomes, expenditure per student is converted to the average over the previous three school years, and then log-transformed. The average number of times city education board member(s) visited school in the city during the previous year, a proxy for school inspection, was 0.73. Autonomy in budget allocation is a dummy variable, which takes the value of one if schools were granted (any degree of) autonomy by the city education board in the allocation of the school budget in the previous year, otherwise takes zero. The mean of 0.25 means that one in four students in the sample attended schools that were given budget autonomy.

For the sake of comparison, we report the national mean of each variable in the last column of Table 1. Note that all of the standardized variables are constructed using all observations from all over Japan. This standardization enables us to identify the impact of information disclosure on test scores using an identification strategy similar to a differences-in-differences approach.

6 Results

We first present results on the unconditional effects on the student's NAAA scores of school accountability via the information disclosure of school-level results of the previous NAAA. We then turn to our main analysis of the effects conditional on the number of neighboring schools.

6.1 Effect of the information disclosure on the student's score

Table 2 presents estimates of the *unconditional* effects both on the student's score in Japanese (Basics) and Total score of the information disclosure of the school-level results of the NAAA. As noted, Japanese (Basics) is the subject component whose school-level score was actually disclosed in September 2013 only in Prefecture A. We sequentially add controls towards Column 4. Specifically, after Column 1 shows the estimate of the effect of information disclosure by regress-

Table 2: Unconditional effects of accountability shock on test scores

	(1)	(2)	(3)	(4)
Dependent variable: Japanese (Basics)				
Treat	2.329***	2.679***	2.949***	1.374***
	(5.481)	(8.170)	(7.206)	(4.088)
Dependent variable: Total				
Treat	1.595***	1.811***	2.013***	1.465***
	(6.956)	(9.672)	(12.832)	(12.926)
Student environment	No	Yes	Yes	Yes
School environment	No	Yes	Yes	Yes
City exp & institution	No	Yes	Yes	Yes
School fixed effects	No	No	Yes	Yes
Lagged school ave score	No	No	No	Yes
Observations		60,917		

Notes: Estimated model applicable to Column (4) is Eq.10 without the interaction terms between the treatment dummy and conditional factors including the number of neighboring schools: $Y_{ism,t} = \beta_1 + \beta_2 Treat_t + \beta_5 \overline{Y}_{sm,t-1} + X_{ismt}\,\eta + Z_{smt}\,\theta + C_{mt}\,\kappa + \upsilon_s + \varepsilon_{ismt}$. The sample period is 2013 and 2014. Dependent variable, $Y_{ism,t}$ is test score standardized within subject and year at the national level with the mean of 50 and the standard deviation of 10. Both Japanese (Basics) score and Total score are considered. T_m is a treatment dummy taking the value of one if $t = 2014$, the year when the first NAAA was implemented after the accountability shock in September 2013. $X_{ism,t}$ ($Z_{sm,t}$, $C_{m,t}$) contains student-level (school-level, city-level education expenditure and institution) variables. υ_s is school fixed effects. Robust t-statistics in parentheses. Clustered standard errors are used to adjust for correlation of error terms within school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

ing without control, Column 2 adds student-, school-, and city-level control variables, and Column 3 also adds school-fixed effects. Last, Column 4 further controls for lagged school-mean scores, corresponding to the model of Eq.10 without the interaction terms between the treatment dummy and conditional factors including the number of neighboring schools.

Coefficients on the treatment dummy are always positive and statistically significant at 1 percent level across the models, regardless of whether the student's score in Japanese (Basics) or Total score is used as a dependent variable. This suggests that the positive effects of the information disclosure are robust to the inclusion of a battery of controls. When the lagged-mean score is taken into account to address the possible mean-reversion in Column 4, the magnitude becomes smaller than Column 3 (particularly for Japanese, Basics). The coefficient of 1.37 for Japanese (Basics) implies that the information disclosure increased the test score of 6th-grade students of treated public primary schools by 1.37 points (i.e., 0.137 standard deviation), compared to those of control schools. According to the model of Column 4, the effect is almost the same when Total score is a dependent variable, except that the standard error (clustered at school level) is much smaller with Total score. Overall, the results on the unconditional effect on student achievement of school accountability induced by the information disclosure is firmly consistent with results from the prior literature (as reviewed in Section 1).[24]

6.2 Role of neighboring schools

We now examine the role of neighboring schools in the information disclosure effect on student achievement. As motivated by the theory in Section 2, the conjecture is that the more neighboring schools there are, the greater the effect of the information disclosure on the student's test score is. The coefficient of main interest is the one on the interaction term between the treatment dummy and the number of neighboring schools, i.e., β_3 in Eq.10. We consider the number of neighboring schools within the 1.5 km radius as a reference. In what follows, all the student-, school-, and city-level variables are controlled for routinely (as in Column (4) of Table 2).

Table 3 presents the estimates of the interaction coefficient when the dependent variable is the student's score in Japanese (Basic). Column (1) estimates the interaction coefficient, β_3, while taking account of the lagged initial score as an additional conditional factor of the information disclosure effect. The interaction coefficient is significantly positive, indicating that the information effect on the test score is greater when there are more neighboring schools. In terms of the magnitude, the coefficient of 0.36 means that an increase in neighboring schools by one increases the

[24]To note, the prior literature indicates that certain ways of information disclosure of school-level performances in a national standardized test can potentially have adverse distributional consequences. For instance, Reback (2008) and Neal and Schanzenbach (2010) find that when school performance measures such as students' pass rates or proportion above some proficiency thresholds are published in the US, schools tend to reallocate resources towards specific students who are critical to improve those measures (albeit in the context of a national standardized test with explicit consequences unlike implicit consequences considered in this paper; see Section 3). In the present context, disclosure of school performances is such that the public was effectively informed of whether or not a school-mean score was above or below the national average (in Japanese, Basics) of the 2013 NAAA. This means that schools had a clear incentive to improve the school average itself. Acknowledging this, however, under such circumstances it is not necessarily clear whether a school resorts to resource reallocation among the students (to improve the average score). While it is possible to conjecture that a school may focus on improving scores of certain students along the distribution, in our companion paper of Morozumi and Tanaka (2020) we showed that when the published school performance measure of a national standardized test (particularly a national assessment) is school-mean scores, such disclosure improves student achievements along the *entire* distribution, without an adverse distributional impact.

Table 3: Role of neighboring schools in the information disclosure effect: Japanese (Basics)

	All within 1.5km radius		Adjacent schools only	
	(1)	(2)	(3)	(4)
Treat	35.026***	37.131***	34.857***	36.814***
	(9.552)	(12.875)	(9.338)	(12.552)
Number of Neighboring sch*Treat	0.360***	0.412***	0.392***	0.440***
	(3.746)	(5.810)	(4.003)	(6.002)
Lagged score*Treat	-0.697***	-0.752***	-0.694***	-0.745***
	(-9.006)	(-12.412)	(-8.805)	(-12.066)
Total number of students*Treat		0.084**		0.078*
		(2.129)		(1.999)
Students needing financial support*Treat		-0.055**		-0.054**
		(-2.211)		(-2.115)
Students needing language help*Treat		-0.173**		-0.170**
		(-2.724)		(-2.783)
School principal experience*Treat		0.516**		0.503*
		(2.110)		(2.064)
Budget autonomy*Treat		0.327		0.335
		(0.884)		(0.904)
All the controls	Yes			
Observations	60,917			

Notes: Estimated model in Columns (1) and (3) are Eq.10: $Y_{ism,t} = \beta_1 + \beta_2 Treat_t + \beta_3 Treat_t * N_{sm} + \beta_4 Treat_t * \overline{Y}_{sm,t-1} + \beta_5 \overline{Y}_{sm,t-1} + X_{ism,t}\,\eta + Z_{sm,t}\,\theta + C_{m,t}\,\kappa + \upsilon_s + \varepsilon_{ism,t}$. The sample period is 2013 and 2014. Model of Columns (2) and (4) further control for interaction terms between the treatment dummy and various possible conditional factors. Dependent variable, $Y_{ism,t}$ is test score standardized within subject and year at the national level with the mean of 50 and the standard deviation of 10. Japanese (Basics) score is considered here. T_m is a treatment dummy taking the value of one if $t = 2014$, the year when the first NAAA was implemented after the accountability shock in September 2013. $X_{ism,t}$ ($Z_{sm,t}, C_{m,t}$) contains student-level (school-level, city-level education expenditure and institution) variables. υ_s is school fixed effects. Robust t-statistics in parentheses. Clustered standard errors are used to adjust for correlation of error terms within school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

effect by 0.036 standard deviation. Incidentally, the interaction between the treatment dummy and the school's lagged score is negative as expected, indicating that when the initial score is lower, the school reacts to the information disclosure more strongly. This result is consistent with findings of the prior literature (see Section 1).

Column (2) controls for more interactions with possible conditional factors. For example, when the total number of students is larger, the effect of the information disclosure is greater. This may be because when the school size is larger, there is more resource available to react to the increase in accountability. The proportion of students who need financial support, an inverse measure of parents' income level, also appears to matter as a conditional factor. The negative and significant coefficient suggests that when the proportion is lower (thus parents' income level is higher at the school level), the effect of the information disclosure is greater, which could be because richer parents, possibly with higher education background, may react to the disclosure more strongly and increase their monitoring of schools more. The negative and significant coefficient of the interaction with the proportion of students who need language help may be interpretable in a similar manner, in that when students have difficulty with understanding Japanese language, parents, who are also likely to have the same problem (possibly because they only recently migrated to Japan), might find it difficult to increase their monitoring level. The positive interaction coefficient with the school principal's tenure (i.e., experience at the school) is also intuitive. When the principal is at the school longer (more than one year, by construction of the dummy), they tend to be more familiar with how the school works, so that they may be in a better position to respond to the accountability shock, being able to utilize the given resource better. Last, the interaction with school budget autonomy is positive, yet not significant. The sign is as expected, since budget autonomy gives the school greater financial discretion when reacting to the information disclosure. The insignificant coefficient may be due to the fact that the variation of this variable is small.[25]

In Columns (3) and (4), we change the way to count the number of neighboring schools. Specifically, while we still consider schools within the 1.5km radius, we here count only adjacent schools, i.e., schools located in the adjacent school districts (instead of all the schools within the radius, as in Columns (1) and (2)). The motivation here is that when neighboring schools are not physically next to each other, parents from those schools may know relatively little about each other even when there is not a long distance between the schools. The results are entirely robust, and in fact, the coefficients on the interaction between the treatment dummy and the number of neighboring schools are now somewhat greater. Moreover, Table A2 in Appendix B shows that the results on

[25] There are two factors that reduce the variation. First, the budget autonomy dummy variable takes the same value across schools within the same city. Second, the time variation of this institutional variable is rather small, such that the fact that we control for school fixed effects leaves little variation exploitable.

the role of neighboring schools in the information disclosure effects are robust when the dependent variable is Total score (the average of Japanese and Mathematics) instead.

Overall, the results on the role of neighboring schools are consistent with the indication from the theoretical model above. Our interpretation of the results is thus as follows. To the extent that external comparison of the standardized test results with neighboring schools (that parents typically know about) facilitates parents' monitoring of the school's effort more than comparison with distant schools, the greater number of neighboring schools induces more monitoring of the school from parents, thus mitigating the moral hazard problem better. Accordingly, schools are incentivized to put more effort (possibly to avoid reputation loss), and the student's score in the national assessment improves further.[26]

6.3 More robustness tests

We now use different radius sizes (other than 1.5 km) to count the number of neighboring schools. Table 4 summarizes estimates of the interaction coefficient between the treatment dummy and the number of neighboring schools for radiuses of 1.3, 1.4, ..., 1.7 km. The upper (lower) half of the table shows results when the dependent variable is the student's score in Japanese, Basics (Total score). As in Table 3, results also cover the case when only adjacent neighboring schools are counted. In all the regressions, we add all the (student-, school-, and city-level) controls, and also the interactions between the treatment dummy and the possible conditional factors (as in Columns (2) and (4) of Table 3). Results are robust to the changes in a radius. Moreover, the coefficients tend to become smaller when the radius becomes larger. For instance, for all the cases (with the different dependent variables and the different ways to count neighboring schools), the interaction coefficients are the smallest when the radius length is the longest. The indication is that the further away schools are located, the less influence they have on the information disclosure effect on the student's test score.

[26] As noted in footnote 12, Burgess et al. (2013) regard the number of neighboring schools (within 5 km) as a measure of local competition. However, since a system of school choice did not operate in the treated prefecture particularly at the primary education level, and the overwhelming majority of primary schools there were public without market incentives, local competition among schools was deemed rather absent. This is why we believe that it is plausible to focus on the role of neighboring schools as a possible mitigating factor of the information problem.

Table 4: Using different radiuses: around 1.5km

Radius	**1.3km**	**1.4km**	**1.5km**	**1.6km**	**1.7km**
	(1)	(2)	(3)	(4)	(5)
Dependent variable: Japanese (Basics)					
All within the radius					
No of Neighbors*Treat	0.399***	0.365***	0.412***	0.343***	0.279***
	(3.799)	(4.755)	(5.810)	(6.228)	(4.789)
Adjacent schools only					
No of Neighbors*Treat	0.450***	0.398***	0.440***	0.387***	0.370***
	(4.204)	(5.091)	(6.002)	(6.038)	(4.980)
Dependent variable: Total					
All within the radius					
No of Neighbors*Treat	0.301***	0.263***	0.299***	0.243***	0.190***
	(2.932)	(3.554)	(4.544)	(4.632)	(3.699)
Adjacent schools only					
No of Neighbors*Treat	0.340***	0.288***	0.321***	0.276***	0.263***
	(3.287)	(4.057)	(4.837)	(4.819)	(4.230)
Other interactions			Yes		
All the controls			Yes		
Observations			60,917		

Notes: Estimated model is a variant of Eq.10 (i.e., with interaction terms between the treatment dummy and all the possible conditional factors considered in Table 3). The sample period is 2013 and 2014. See Notes for Table 3 for more explanation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To shed further light on the relevance of *closeness* between the school and neighboring schools, we compare the relevance (to the information disclosure effect) of the number of schools "within the radius of 1.5 km" with the number of schools "within the 1.5 - 3km annulus area" (i.e., the annulus area between the concentric circles with the radiuses of 1.5 and 3 km) and the number of schools "within the 3 - 4.5 km annuls area". To the extent that it is the closeness to the neighboring school that matters, we should find that the number of schools within the further annulus area should be less relevant to the information disclosure effect. Table 5 shows that this is exactly the case. Specifically, in the upper half of the table (where the dependent variable is Japanese, Basics), Column (1) finds that the coefficient on the interaction between the treatment dummy and the number of neighboring schools within the 1.5 km radius is significantly positive with the magnitude of 0.41 (as in Column (2) of Table 3). Turning to the results on the numbers of

Table 5: Consider number of schools within annuluses: within the circle of 4.5km

Annulus with radius of	**0-1.5km**	**1.5-3km**	**3-4.5km**
	(1)	(2)	(3)
Dependent variable: Japanese (Basics)			
All within the radius			
No of Neighbors*Treat	0.412***	0.126***	0.009
	(5.810)	(3.103)	(0.303)
Dependent variable: Total			
All within the radius			
No of Neighbors*Treat	0.299***	0.092**	0.017
	(4.544)	(2.641)	(0.783)
Other interactions		Yes	
All the controls		Yes	
Observations		60,917	

Notes: Estimated model is a variant of Eq.10 (i.e., with interaction terms between the treatment dummy and all the possible conditional factors considered in Table 3). The sample period is 2013 and 2014. See Notes for Table 3 for more explanation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

schools within the 1.5 - 3 km and 3 - 4.5 km annulus area, Column (2) and (3) show that as we consider the number of schools further away, the relevance fades gradually; the number of schools within the former annulus area makes a quantitatively much smaller impact, albeit statistically significant, whereas the number within the latter annulus area becomes entirely irrelevant. The same conclusion is drawn when the dependent variable is Total score. The clear indication is thus that it is *close* neighboring schools that are relevant to the information disclosure effect.

7 Discussions: Are neighboring schools always relevant?

Having established the relevance of close neighboring schools to the information disclosure effects, we explore the further dimension of heterogeneity. Specifically, we now consider the possibility that the number of neighboring schools is relevant to the effects *only when* schools were revealed as a low-performer and thus presumably under stronger accountability pressure. Remember that in the accountability shock in September 2013 that constitutes a natural experiment, the then governor published the names of school principals whose school-mean scores in Japanese (Basics) were above the national-mean score in that subject, effectively categorizing all the primary schools in

Table 6: Disclosed as low- (high-) performing schools in 2013

	Adjacent schools only within 1.5km radius		
Sample	**All**	**Low (Below national ave)**	**High (Above national ave)**
	(1)	(2)	(3)
Dependent variable: Japanese (Basics)			
No of Neighbors*Treat	0.440***	0.467***	0.123
	(6.002)	(5.857)	(0.866)
Dependent variable: Total			
No of Neighbors*Treat	0.321***	0.358***	-0.020
	(4.837)	(4.860)	(-0.178)
Other interactions		Yes	
All the controls		Yes	
Observations	60,917	51,958	8,959

Notes: Estimated model is a variant of Eq.10 (i.e., with interaction terms between the treatment dummy and all the possible conditional factors considered in Table 3). The sample period is 2013 and 2014. Columns (2) and (3) consider students attending schools whose mean score of Japanese (Basics) was below and above the national mean of the subject in 2013, respectively. See Notes for Table 3 for more explanation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the prefecture into the two groups: ones that scored above and below the national mean in the 2013 NAAA. The idea now is to divide the sample into the two sub-samples following this categorization, and to estimate whether the number of neighboring schools has the same relevance to the information disclosure effect across the two sub-samples.

Table 6 summarizes the estimates of the coefficient on the interaction between the treatment dummy and the number of neighboring schools (adjacent schools in particular) within the 1.5 km radius, for the two sub-samples. Column (1) replicates the result of Column (4) of Table 3 for the whole sample, and Columns (2) and (3) show results when the sub-sample only contains students from the schools whose mean score (in Japanese, Basics) was below and above the national mean score in 2013, respectively. As observed, the number of neighboring school is a significant conditioning factor of the information disclosure effect only when schools were revealed as a low-performer in the 2013 information disclosure. This suggests that external comparability of the results, particularly the comparability with neighboring schools, only affects the information disclosure effects when parents recognize the school as a low-performer. This appears intuitive, because if the school is exposed as a high-performer, parents may not have an incentive to monitor the school's effort level closely to begin with.

8 Concluding Remarks

We showed that while the disclosure of school-level results of a national assessment has a potential to improve student achievements, the effects are heterogeneous across schools. In particular, the information disclosure effect critically depends on the number of neighboring schools. The principal agent model suggests that when it is external comparison with neighboring schools (rather than distant schools) that helps parents monitor the school's effort, the greater number of neighboring schools mitigates the moral hazard problem more, thereby incentivizing the school to increase the effort and improve student achievements more. Consistent with the theory, the empirical analysis, utilizing the accountability shock that happened only in one of the prefectures in Japan in 2013, showed that the information disclosure does cause a significantly greater effect on student achievements when the school has more neighboring schools. The results further suggested that neighboring schools matter particularly when the school was revealed as a low-performing school.

This paper contributes to the growing literature on the information disclosure effects of school-level performances in a national standardized test. The fact that a national standardized assessment is enormously popular all over the world (e.g., OECD (2015) and World Bank (2018)) only reinforces the importance of this paper's findings. While it has been known that the level of aggregation at which results of such a test is disclosed is a critical design feature of the test, our results shed new light on the condition under which disclosure of the results at the school level has a particularly desirable effect on student outcomes.

Data Availability Statement: This paper uses confidential data from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Researchers may apply for the secondary use of the data and can use them if research proposals are approved by the Institutional Review Board of the Ministry. The information of application process is available from the Ministry's webpage: https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/sonota/1386492.htm

Disclosure Statement: This study is conducted as a part of the Project "Microeconometric Analysis of Education Policy with Large Administrative Data" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). Morozumi declares that he has no relevant or material financial interests that relate to the research described in this paper. Tanaka declares that his research is supported by the Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (grant number: 20H05629). This research has been approved by the Institutional Review Board of the Ministry of Education, Culture, Sports, Science and Technology of Japan. There is no other conflict of interest to be disclosed for all the authors.

Appendix

A Data Sources

Table A1: Data sources

Data (alphabetical order)	Sources
National assessment scores	National Assessment of Academic Ability (NAAA)
Number of primary school students at the city level[a]	e-Stat (Portal site for Japanese Government Statistics) https://www.e-stat.go.jp/en
Number of schools in close proximity[b]	NAAA & Google Maps
Other school environment variables (see Table 1)	NAAA
Parental school choice	Our own survey and interview with city educ boards
Primary education expenditures at city level[c]	e-Stat
School budget autonomy	Survey on Education Boards
School inspection by educ boards	Survey on Education Boards
School principal's tenure	National School Directory
Student characteristics variables (see Table 1)	NAAA

Notes: (a) Used to obtain primary education expenditure per primary school student at the city level. (b) Number of (primary) schools within the 1.5 km radius in the reference analysis. (c) Divided by the number of primary school students in the city, and averaged over the past three years (to match spending at the time when students were in 3rd, 4th, and 5th grades).

B Supplementary results

Table A2: Role of neighboring schools in the information disclosure effect: Total score

	All within 1.5km radius		Adjacent schools only	
	(1)	(2)	(3)	(4)
Treat	24.168***	26.796***	23.987***	26.446***
	(9.410)	(9.835)	(9.614)	(9.997)
Number of Neighboring sch*Treat	0.255***	0.299***	0.278***	0.321***
	(3.699)	(4.544)	(4.046)	(4.837)
Lagged score*Treat	-0.468***	-0.521***	-0.464***	-0.514***
	(-8.768)	(-9.312)	(-8.968)	(-9.460)
Total number of students*Treat		0.030		0.025
		(1.019)		(0.867)
Students needing financial support*Treat		-0.044**		-0.043**
		(-2.516)		(-2.505)
Students needing language help*Treat		-0.088*		-0.086*
		(-1.799)		(-1.797)
School principal experience*Treat		0.062		0.047
		(0.339)		(0.259)
Budget autonomy*Treat		0.031		0.036
		(0.152)		(0.178)
All the controls	Yes			
Observations	60,917			

Notes: Estimated model in Columns (1) and (3) are Eq.10: $Y_{ism,t} = \beta_1 + \beta_2 Treat_t + \beta_3 Treat_t * N_{sm} + \beta_4 Treat_t * \overline{Y}_{sm,t-1} + \beta_5 \overline{Y}_{sm,t-1} + X_{ism,t}\eta + Z_{sm,t}\theta + C_{m,t}\kappa + \upsilon_s + \varepsilon_{ism,t}$. Model of Columns (2) and (4) further control for interaction terms between the treatment dummy and various possible conditional factors. Dependent variable, $Y_{ism,t}$ is test score standardized within subject and year at the national level with the mean of 50 and the standard deviation of 10. Total score is considered here. See Notes for Table 3 for more explanation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

References

BERGBAUER, A. B., E. A. HANUSHEK, AND L. WOESSMANN (2021): "Testing," Journal of Human Resources, October 13, 2021 doi: 10.3368/jhr.0520-10886R1.

BISHOP, J. (2006): "Drinking from the fountain of knowledge: Student incentive to study and learn - Externalities, information problems and peer pressure," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek and F. Welch, Elsevier, vol. 2, 909–944.

BURGESS, S., D. WILSON, AND J. WORTH (2013): "A natural experiment in school accountability: The impact of school performance information on pupil progress," *Journal of Public Economics*, 106, 57–67.

CAMARGO, B., R. CAMELO, S. FIRPO, AND V. PONCZEK (2018): "Information, market incentives, and student performance: Evidence from a regression discontinuity design in Brazil," *Journal of Human Resources*, 53, 414–444.

CHIANG, H. (2009): "How accountability pressure on failing schools affects student achievement," *Journal of Public Economics*, 93, 1045–1057.

CILLIERS, J., I. M. MBITI, AND A. ZEITLIN (2021): "Can public rankings improve school performance?" *Journal of Human Resources*, 56, 655–685.

FERREYRA, M. AND P. LIANG (2012): "Information asymmetry and equilibrium monitoring in education," *Journal of Public Economics*, 96, 237–254.

FIGLIO, D. AND S. LOEB (2011): "School Accountability," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Amsterdam: Elsevier, vol. 3, 383–421.

FIGLIO, D. N. AND H. F. LADD (2007): "School accountability and student achievement," in *Handbook of Research in Education Finance and Policy*, ed. by H. F. Ladd and E. B. Fiske, London: Routledge, 166–182.

HANUSHEK, E. A., S. LINK, AND L. WOESSMANN (2013): "Does school autonomy make sense everywhere? Panel estimates from PISA," *Journal of Development Economics*, 104, 212–232.

HANUSHEK, E. A. AND M. E. RAYMOND (2005): "Does school accountability lead to improved student performance?" *Journal of policy analysis and management*, 24, 297–327.

KONING, P. AND K. VAN DER WIEL (2012): "School responsiveness to quality rankings: An empirical analysis of secondary education in the Netherlands," *De Economist*, 160, 339–355.

MOROZUMI, A. AND R. TANAKA (2020): "Should School-Level Results of National Assessments Be Made Public?" IZA, Discussion Paper 13450.

NEAL, D. AND D. W. SCHANZENBACH (2010): "Left behind by design: Proficiency counts and test-based accountability," *Review of Economics and Statistics*, 92, 263–283.

OECD (2011): "School autonomy and accountability: Are they related to student performance?" PISA in Focus, No.9, OECD Publishing, Paris, https://doi.org/10.1787/5k9h362kcx9w-en.

——— (2015): "Education at a Glance 2015: OECD Indicators," OECD Publishing, Paris, http://dx.doi.org/10.1787/eag-2015-en.

REBACK, R. (2008): "Teaching to the rating: School accountability and the distribution of student achievement," *Journal of Public Economics*, 92, 1394–1415.

ROCKOFF, J. AND L. J. TURNER (2010): "Short-run impacts of accountability on school quality," *American Economic Journal: Economic Policy*, 2, 119–147.

ROUSE, C. E., J. HANNAWAY, D. GOLDHABER, AND F. DAVID (2013): "Feeling the Florida Heat? How low-performing schools respond to voucher and accountability pressure," *American Economic Journal: Economic Policy*, 5, 251–281.

WORLD BANK (2018): *World Development Report 2018: Leaning to realize education's promise*, Washington, DC: World Bank.