Children and Youth Services Review 113 (2020) 104896

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth

Red flag: Grade retention and student academic and behavioral outcomes in China

Li-Chung Hu[a,*], Emily Hannum[b]

[a] Department of Sociology, National Chengchi University, Taipei, Taiwan, ROC
[b] Department of Sociology, University of Pennsylvania, Philadelphia, USA

ARTICLE INFO

Keywords:
Grade repetition
propensity score matching
difference-in-difference
externalizing behavioral problems
school transfer
parental migration

ABSTRACT

This paper analyzes the prevalence, correlates, and behavioral and academic impacts of grade retention using national and single-province data from China. Retention is a more common experience in China than official estimates suggest; it is more frequent in less-developed parts of the country; and it is associated nationally with poorer subsequent performance and psychosocial well-being, even after adjusting for numerous confounders. However, with certain caveats, findings suggest that retention is primarily a "red flag" and is not a cause of poorer achievement and behavioral outcomes. A longitudinal analysis in one province shows that retained children can gain ground in academic and behavioral outcomes; a further causal analysis using matched samples and difference-in-difference approaches shows no evidence of a causal impact of retention on outcomes. High levels of population mobility and associated school transfers may contribute to grade retention being reported by students and families but not captured in school records.

1. Introduction

The practice of grade retention or grade repetition occurs in both developed and developing countries. Economically disadvantaged children may be particularly affected: retention tends to be more common among socioeconomically disadvantaged students (OECD, 2016). In China, the largest educational system in the world, official estimates indicate that retention rates are miniscule. However, the national scale of grade retention in China is not well established, and regional surveys indicate that a sizeable minority of children do repeat a grade (for example, see Chen, Liu, Zhang, Shi, & Rozelle, 2010; Liu, Zhang, Luo, Rozelle, & Loyalka, 2010; OECD, 2016). Globally, evidence about the impact of grade retention on academic and behavioral outcomes is decidedly mixed (Ikeda & García, 2014, 2012). In China, the impact of grade retention on student outcomes has been little studied.

This paper begins to address the lack of attention to retention in China, first by describing the national scale of primary grade retention and its associations with subsequent outcomes, and second by performing an impact evaluation of the academic and behavioral consequences of retention in poor rural communities in one province. Specifically, we first investigate two questions with national data: 1. Are students in less well-off regions of China more vulnerable to grade retention? 2. Do associations of retention with academic and psychosocial outcomes differ in more rural, isolated areas, compared to wealthier urban areas? Here, we utilize cross-sectional data from the baseline wave of the China Education Panel Study (CEPS), a nationally representative middle school student survey, to establish the scope of primary retention. We then focus on poor rural communities in one province to address a third question: 3. Is there an estimable causal impact of retention on academic and behavioral outcomes? In these analyses, we apply difference-in-difference and matching approaches to data from the Gansu Survey of Children and Families (GSCF), which contains detailed records of grade repetition, long-term educational outcomes, and potential confounders for a sample of rural children in Gansu Province.

2. Background

2.1. Grade retention in comparative perspective

Grade retention occurs around the globe in both developed and developing countries (Hungi, 2010, 2009), and may be a more common experience in poor areas and among socioeconomically disadvantaged students (OECD, 2016). Comparative research on grade retention has suggested that it is associated with negative academic and behavioral outcomes in many but not all settings. In the United States, first grade repetition rates for 2008–2009 were about 3.5 percent (Warren & Saliba, 2012). There, grade retention is often viewed as a risk factor for

* Corresponding author at: Department of Sociology, National Chengchi University, No. 64, Sec. 2, Zhinan Rd., Wenshan District, Taipei City 11605, Taiwan, ROC.
E-mail address: lchu@nccu.edu.tw (L.-C. Hu).

https://doi.org/10.1016/j.childyouth.2020.104896
Received 29 October 2019; Received in revised form 24 February 2020; Accepted 24 February 2020
Available online 25 February 2020
0190-7409/

school dropout (Jimerson, Anderson, & Whipple, 2002). In a comprehensive literature review, Holmes (1989) evaluated the association between grade retention, educational performance, and children's psychosocial wellbeing, and concluded that there was no meaningful benefit associated with grade retention. On the contrary, retention was detrimentally associated with psychosocial wellbeing and welfare outcomes. Jimerson (2001) reviewed updated research from 1990 to 1999 and drew a similar conclusion: the majority of existing studies suggest that grade retention is harmful or not beneficial to academic achievement and socioemotional and behavioral adjustment, with only a handful of studies suggesting otherwise.

Grade retention is rated as one of the most stressful events by students. One study suggests that sixth grade students in the United State rate the stressfulness of being retained as greater than even the death of a parent or going blind (Anderson, Jimerson, & Whipple, 2005). Scholars theorize that grade retention can act as a negative trigger, with scarring implications for subsequent behavioral adjustment, psychosocial development, and educational achievement and attainment (Andrew, 2014; Goos et al., 2013; Pagani et al., 2001). From a life course perspective, grade retention at early stages of education may be a turning point that potentially triggers children's delinquent behavior, harms their self-esteem, detracts from their school engagement, and leads to early school dropout (Alexander, Entwisle, & Kabbani, 2001; Audas & Willms, 2002; Nagin, Pagani, Tremblay, & Vitaro, 2003; Stearns, Moller, Blau, & Potochnick, 2007).

Cross-national studies have yielded similar findings, though with some exceptions. In an analysis of data published by PISA in 2009, Ikeda and García (2014) examine the relationship between grade retention and academic and non-academic outcomes in 30 countries (including Hong Kong and Macau). The authors find negative associations between grade retention and reading performance in 21 of them. Regarding non-academic outcomes (measured by students' attitudes toward school), non-repeaters tend to report more positive attitudes toward school in fewer than half (13) of these countries. The authors argue that cross-national variations in the relationship between grade retention and academic and non-academic outcomes can be attributed to how repeaters are treated in different education systems. Repeating a grade during primary school can be viewed as a learning strategy in places where educational resources are scarce – a phenomenon not uncommon in developing countries, particularly in rural and remote areas (Brophy, 2006). In areas with more educational resources, however, repetition may be utilized differently or less frequently.

Studying the impact of grade retention on subsequent academic and non-academic outcomes faces at least two significant challenges: adequately addressing selection and adequately addressing context. On the first point, the selectivity of retention is poorly addressed in many studies, either due to lack of important pre-retention covariates or reliance on conventional statistical controls (Allen, Chen, Willson, & Hughes, 2009; Lorence, 2006). Researchers have adopted various analytical strategies to address potential retention selection issues based on observed or unobserved covariates. Jacob and Lefgren (2004) apply a regression-discontinuity design to evaluate the impact of primary grade retention on math and reading scores in Chicago public schools. They find no evidence to support the hypothesis that grade retention hinders children's educational achievement. On the contrary, they find that retention improves educational achievement in low-achieving students. Using the same analytical strategy, Jacob and Lefgren (2009) evaluate the influence of grade retention on high school completion. Their findings, namely that the negative impact of grade retention on high school completion is rather modest, challenge the popular view that grade retention impedes high school completion in the United States. Recent meta-analysis also suggests that the negative impact of grade retention tends to dissipate when the selectivity of retention is adequately addressed (Allen et al., 2009).

Second, exposure to and consequences of grade retention may be context-specific. For example, when passing standardized exams is required for advancing to the next grade, or when accountability is emphasized, retention rates may be higher. In the United States, in several states, the number of students being retained increased after the incorporation of standards-based testing as a central determinant of grade advancement (Bali, Anagnostopoulos, & Roberts, 2005; Warren & Saliba, 2012). To provide another example, using nationally representative data and utilizing propensity score matching and sibling-fixed effects estimation, Andrew (2014) contradicts the study of Jacob and Lefgren (2009) and shows that grade retention does in fact negatively impact educational transitions. The author argues that although the Chicago public schools studied by Jacob and Lefgren had higher retention rates than the national average, summer tutoring was offered to help low-achieving students. Thus, the causal impact of grade retention may have differed in that setting.

A key element of context has to do with educational resources. Schools in high-resource communities may not need to resort to retention very frequently and may also be able to draw on resources such that students receive greater academic support when retained. At the same time, if stigma is perceived to be an important mechanism through which grade retention shapes subsequent academic and non-academic outcomes, it would be reasonable to suspect that the negative psychological impact of retention may be smaller in contexts—often lower resource contexts—with high retention rates (Demanet & Van Houtte, 2013; Hong and Raudenbush 2005, 2006). Thus, schools in low-resource communities may be more likely to utilize retention as a strategy, but less likely to provide strong resources to support academic development of retained students. Further, retained students in low-resource schools where repetition is a more common experience may be less stigmatized by the experience. Overall, these studies suggest that both vulnerability to and the impact of grade retention may differ in wealthier and poorer regions.

2.2. Grade retention in China

In the past 30 years, China has dramatically improved access to education. In the context of significant urban-rural and regional economic disparities, the State has largely eliminated barriers to primary and lower secondary education and expanded access to higher education while maintaining an extremely competitive high-stakes testing system (Hannum, 1999; Wan, 2006; Yu & Suen, 2005). Given the high-stakes testing regime, one might expect to find high rates of retention. However, according to official estimates, this is not the case: not only is the percentage of repeaters in primary school extremely low, but it has even been decreasing, having dropped from 1.5 percent in 2006 to 0.1 percent by 2014[1] (UNESCO Institute for Statistics, 2018). This low official rate may be due to the government's discouragement of retention and regulation of the maximum proportion of repeaters in each school (Chen et al., 2010).

However, survey-based estimates in some contexts suggest that the prevalence of children repeating a grade during compulsory education may be higher than suggested by official, school-reported estimates. For instance, recent studies estimate that about 30 percent of children in Shaanxi Province repeated a grade during primary school (Chen et al., 2010). Analysis of PISA data from four regions in China indicates that the overall percentage of students who report having repeated a grade in primary or secondary school is 20.8 percent,[2] but rates differ by

[1] The definition of percentage of repeaters in primary education in the official data is the "total number of students who are enrolled in the same grade as the previous year, expressed as a percentage of total enrollment in the given grade of education" (UNESCO Institute for Statistics, 2018). The official data is a period data and is not directly comparable to cohort data. These numbers are estimated based on the official data in order to compare with survey data.

[2] In PISA, 15-year-old students reported whether they repeated a grade in at least one International Standard Classification of Education (ISCED) level. Thus, the prevalence of grade retention in PISA includes both primary and lower

socioeconomic group: nearly 30 percent of socioeconomically disadvantaged children report retention, while just under 10 percent of advantaged students do (OECD, 2016).

Some of the discrepancy between what is captured in official statistics and what children and families report may have to do with how retention happens. Repetition could be initiated by the school, but it is possible that some parents opt for retention out of concern that their children may not be academically prepared to advance to the next grade. Parent-initiated retention may not always get recorded in official statistics. It can happen when parents move their children to a different school in consideration of the children's psychosocial health, or for other reasons such as school closures and the circumstances of the parents' employment.[3] These latter two reasons may have become more significant in recent years due to the massive increase of migrant laborers in cities across China and the resulting school closures in rural areas with dwindling populations. In these cases, it is possible that repetition goes unrecorded in official records. This phenomenon may be especially widespread in impoverished rural contexts, where repeating a grade may be the least expensive or only feasible means of helping students who are experiencing academic difficulties in an education system characterized by high-stakes exams and limited resources.

Few studies have investigated the implications of grade retention in China. Chen et al. (2010) evaluate the causal influence of grade retention in Shaanxi Province using a difference-in-difference propensity score matching approach. They find that grade retention has positive effects on grades in math and Chinese. However, the benefits of grade retention disappear when selectivity of grade retention on observed covariates is considered. Examining the linkage between repeating a grade during primary school and educational transitions among rural children using longitudinal data, one study finds that retention was a risk factor associated with non-completion of middle school and non-continuation to comprehensive high schools (Hu & Hannum, in press).

In sum, government statistics suggest minimal officially-recognized grade repetition in China, but numbers reported by students in surveys suggest that the experience is non-trivial. Perhaps due to the low official estimates, few studies have sought to describe the scope of grade retention in China or track its implications. This paper addresses these limitations. It investigates the scope of grade retention across China and tests whether the associations between retention and academic and psychosocial outcomes differ across wealthier urban regions and more rural regions. It then draws on detailed longitudinal data from one province to perform an evaluation of the causal impact of grade retention on academic and psychosocial outcomes by applying difference-in-difference and propensity score matching approaches to address both observed and unobserved selection.

3. Data and methods

We analyze two data sources: a national, cross-sectional dataset to describe patterns and test regional differences and a single-province longitudinal dataset to perform an impact analysis by applying difference-in-difference and propensity score matching approaches.

3.1. The China Education Panel Survey (CEPS)

The CEPS is a representative sample of middle school students in China conducted in 2013–2014 using a multistage stratified random sampling design (National Survey Research Center, 2020). First, probability proportionate to size sampling (PPS) was used to randomly select 28 out of 2870 counties. Four schools in each county were randomly chosen (4*28). Then, two 7th grade classes and two 9th grade classes were randomly selected in each school. The survey was administered to all students, along with their teachers and parents, in the randomly selected classes. Given its representativeness, the CEPS provides a fairly accurate overview of the prevalence of grade retention across China and facilitates an examination of the relationship between grade retention and academic and non-academic outcomes.

We excluded cases with missing values for dependent and continuous independent variables. We further treated missing values for categorical variables as one specific category and included them in the analysis to retain as many cases as possible. Overall, around 92% of cases were maintained in the final analytical sample. Those who were not included in the final sample tended to have lower achievement and performance on tests, more depressive symptoms, and higher retention rates, and they tended to be from more disadvantaged backgrounds. Thus, the negative associations between dependent variables of interest and retention tended to be underestimated.

3.1.1. Measurement

Tables 1a and 1b contain descriptive statistics for the CEPS and GSCF datasets. In the CEPS data, our dependent variables are educational achievement and psychological wellbeing (for more details, please see Appendix A). Educational achievement was measured by testing math, Chinese, English, and cognitive skills.[4] Psychological wellbeing is regarded as an absence of depressive symptoms,[5] as determined by the child's response to the item "Within the past seven days, have you felt dispirited, despondent, unhappy, sorrowful, or that life is meaningless?"[6] Our key analytical variable *grade retention* is child-reported, with each child asked if he or she had been held back between the 1st and 6th grades.

To capture the effects of school transfers, which have been found to be a possible source of disruption and disadvantaged educational performance (OECD, 2011),[7] we include the variable *school transfer*. Past

(*footnote continued*)
secondary education repetition.

[3] According to 2009, the category of repeaters ought to include students who repeat the same grade more than once and students who repeat the same grade because they transferred from one school to another. Factors that may throw off estimates of the prevalence of grade repetition include enrollment over-reporting and the conflation of new entrants, repeaters, and transfer students who repeat the same grade at a new school.

[4] These grades are curriculum-based and standardized within each grade. Because the exam and evaluation criteria are not uniform across schools, the raw scores are not comparable across schools. For this reason, the raw scores were converted to standardized scores by the schools themselves. Teachers were asked to report the sample child's grades in math and language in the previous semester based on the child's transcript. If grade retention is viewed both as a strategy to help low-achieving students and a process of learning and evaluating student performance, improvement in grades should be expected and the grade gap between retained and non-retained students should be narrower. In the CEPS, standardized cognitive tests designed by researchers were administered to every sample child. These tests are curriculum-irrelevant and can be used to compare different schools.

[5] A reviewer made the important point our measure of psychological well-being is narrow and fails to encompass a full range of important constructs, such as the six dimensions proposed by Ryff (1995): autonomy, environmental mastery, personal growth, positive relationships with other people, purpose in life, and self-acceptance. We acknowledge the limitation of our measurement strategy, which is constrained by available data. A broader measure of psychological well-being is certainly needed to fully understand the potentially stigmatizing implications of retention.

[6] Children responded to each item using a 5-point Likert scale, with 1 for "never," 2 for "seldom," 3 for "sometimes," 4 for "often," and 5 for "always." The Cronbach's alpha was 0.86. We created a scale from Cronbach's alpha based on these 5 items.

[7] On the one hand, school transfer may be chosen by parents when a child is retained or expected to be retained, as a possible means of remedying low achievement and behavioral problems (OECD, 2011). Parents may also transfer their children to another school if school and teacher quality are major concerns. On the other hand, disruptions faced by transfer students may become a

research has indicated that school entrance age is an important factor in understanding children's educational achievement, behavioral problems, and grade retention (Bedard & Dhuey, 2006; Martin, 2009; McEwan & Shapiro, 2008). Therefore, we include *age upon entering primary school* in the analysis. Kindergarten attendance was not required by the Chinese government at the time of this study, but parents may have sent their children to kindergartens at their own expense if kindergartens were available in the community. Thus, we include *non-kindergarten attendance* as a measure of disadvantage.[8]

Three other types of independent variables included in the analysis are child characteristics, family background, and school dummies. Child characteristics include *girl*, *only child*, *Han* (the majority ethnic group in China), *age*, *migration status*, *hukou status* (indicating urban or rural residence), and *health problems*,[9] all of which have been recognized as having some associations with educational outcomes or mental health in existing literature (Alderman, Behrman, Lavy, & Menon, 2001; Jamison, 1986; Zeng, Pang, Zhang, Medina, & Rozelle, 2014). *Grade* children attended and *semester* during which the survey was conducted are also included. Various dimensions of family background, including *father's* and *mother's migration status, parental educational expectations*, *father's* and *mother's education*, and *family wealth* (measured in quintiles). *School dummy variables*[10] are also incorporated to account for potential confounding effects.

3.1.2. Methods

To describe the scale and correlates of grade retention as well as the association between grade retention, educational achievement, and psychological health, we use the CEPS data to conduct descriptive analysis and ordinary least squares regression with school dummy variables. We add to this model a series of interaction terms to test regional differences in the association of retention with academic and behavioral outcomes.

3.2. The Gansu Survey of Children and Families

The GSCF is the first longitudinal survey launched in rural China to examine the relationships among child welfare, family dynamics, and social mobility. The GSCF has been utilized to examine how factors such as gender, nutrition, behavioral problems, family separation, parenting, home environment, and poverty relate to educational outcomes (Cherng & Hannum, 2013; Hannum & Hu, 2017; Hannum, Hu, &

Table 1a
Descriptive Statistics for all Variables in the CEPS (Weighted).

	N	Mean	SD	Min	Max
Dependent variables					
Math	18,989	−0.09	1.04	−3.56	2.94
Chinese	18,988	−0.18	1.09	−4.68	2.66
English	18,996	−0.11	1.02	−4.04	2.88
Depressive symptoms	19,091	2.11	0.78	1.00	5.00
Cognitive test	19,487	−0.15	0.85	−2.03	2.71
Independent variable					
Child characteristics					
Gender					
Girl	19,487	0.48	0.50	0	1
Missing	19,487	0.01	0.12	0	1
Only child	19,487	0.33	0.47	0	1
Han	19,487	0.84	0.36	0	1
Child age	19,025	13.72	1.29	11	17
Retention	19,421	0.22	0.42	0	1
Child migration status (ref.: local non-immigrant)					
Intra-provincial	19,366	0.06	0.24	0	1
Inter-provincial	19,366	0.04	0.19	0	1
Agricultural hukou	19,487	0.64	0.48	0	1
Health problems					
Yes	19,487	0.11	0.32	0	1
Missing	19,487	0.05	0.22	0	1
Educational experience					
School transfer	19,403	0.32	0.47	0	1
Age upon entering primary school	19,234	6.42	1.05	3	9
Non-kindergarten attendance					
No	19,487	0.20	0.40	0	1
Missing	19,487	0.01	0.09	0	1
Parents' educational expectations					
College and above	19,487	0.67	0.47	0	1
Missing	19,487	0.04	0.19	0	1
9th grade (ref.: 7th grade)	19,487	0.50	0.50	0	1
Fall semester, 2013 (ref.: spring semester, 2014)	19,487	0.69	0.46	0	1
Family background					
Parental migration status (both parents at home)					
Only mother at home	19,487	0.11	0.31	0	1
Only father at home	19,487	0.04	0.19	0	1
Neither parent at home	19,487	0.14	0.35	0	1
Mother's education (ref.: primary school and less)					
Middle school	19,171	0.42	0.49	0	1
Secondary specialized/technical school	19,171	0.04	0.20	0	1
High school or vocational high school	19,171	0.13	0.33	0	1
Junior college	19,171	0.03	0.18	0	1
College and above	19,171	0.05	0.21	0	1
Father's education (ref.: primary school and less)					
Middle school	19,441	0.49	0.50	0	1
Secondary specialized/technical school	19,441	0.04	0.20	0	1
High school or vocational high school	19,441	0.17	0.37	0	1
Junior college	19,441	0.04	0.19	0	1
College and above	19,441	0.06	0.23	0	1
Family wealth (ref.: poorest)					
Relatively poor	19,487	0.21	0.44	0	1
Average	19,487	0.65	0.49	0	1
Relatively well-off /well-off	19,487	0.04	0.17	0	1
Missing	19,487	0.05	0.21	0	1
School context					
Administrative level (ref.: direct-controlled municipality)					
Provincial capital	19,487	0.09	0.28	0	1
Prefecture-level city	19,487	0.18	0.39	0	1
County	19,487	0.71	0.45	0	1

(*footnote continued*)
risk that can potentially lead to grade retention (Guèvremont, Roos, & Brownell, 2007). Incorporating the effects of school transfer into the analysis is especially relevant to the Chinese educational system, where school choice is pervasive and massive school merger programs dating from 2001 have caused huge numbers of students to change schools (Qin, 2008; Kipnis, 2006; Mei, Jiang, Xiang, & Song, 2015; Tsang, 2003). These two widespread phenomena, both of which may contribute to or stem in part from grade retention, likely contribute to students transferring primary schools in China. Ideally, we would have information about the sequential order of school transfer and grade retention and analyze this data within a longitudinal framework. Unfortunately, only the CEPS data includes school transfer information, and it is impossible to identify whether grade retention occurs before school transfer or vice versa. Thus, we are unable to determine whether school transfer is an outcome or a risk factor of grade retention. Nonetheless, it is certain that school transfer is a potential confounding factor that is correlated with educational achievement and grade retention. Therefore, school transfer is included in the model to control the potential influence of school transfer on the dependent variables of interest, thus providing a conservative estimation of grade retention effects.

[8] In this variable, "0" represents children who attended kindergarten; "1" represents children who did not attend kindergarten; and "2" represents missing values. This variable is further treated as two dichotomous variables in the analysis.

[9] "Health problems" is measured by asking parents whether this child had severe illness before attending primary school. "1" represents "Yes" and "0" represents "No". Missing is included as another category. "0" represents no severe illness; "1" represents had severe illness and "2" represents missing values. This variable is further treated as two dichotomous variables in the analysis.

[10] *School dummy variables* refers to 119 dichotomous, 0–1 variables, which collectively identify all 120 schools in the CEPS. By using school dummy variables, we are able to account for school-specific differences in average level of educational achievement and behavioral problems in the analysis.

Table 1b
Descriptive Statistics for all Variables in the GSCF.

	N	Mean	SD	Min	Max
Dependent variables					
EBP, 2004	1611	−0.01	1.01	−1.97	4.52
IBP, 2004	1611	0.00	1.00	−2.89	4.22
Math, 2004	1184	0.00	0.94	−3.61	2.30
Chinese, 2004	1181	0.00	0.94	−3.18	2.93
Change between 2000 and 2004					
EBP	1611	−0.01	1.37	−5.05	5.46
IBP	1611	0.00	1.41	−6.00	4.65
Math	1184	−0.05	1.23	−5.05	4.34
Chinese	1181	−0.05	1.22	−4.49	3.96
Independent variables					
Child characteristics					
Girl	1611	0.53	0.50	0	1
Only child	1611	0.06	0.23	0	1
Child age	1611	10.92	0.97	9	12
Health problems	1611	0.04	0.19	0	1
Early retention	1611	0.37	0.48	0	1
Late retention	1611	0.22	0.41	0	1
Educational experience					
Age upon entering primary school	1611	7.40	0.94	5	12
Non-kindergarten attendance	1611	0.72	0.45	0	1
EBP, 2000	1611	0.01	1.02	−1.96	4.07
IBP, 2000	1611	0.00	1.01	−2.69	3.88
Math grade, 2000	1611	0.01	0.95	−3.67	2.82
Chinese grade, 2000	1611	0.02	0.95	−4.10	2.79
Family background					
Father or mother are away from home	1611	0.08	0.27	0	1
Mother's educational expectation	1611	11.98	2.84	6	16
Father's education (ref: less than primary school)					
Above primary school and below middle school	1611	0.17	0.38	0	1
Middle school and above	1611	0.46	0.50	0	1
Mother's education (ref.: less than primary school)					
Above primary school and below middle school	1611	0.13	0.34	0	1
Middle school and above	1611	0.19	0.39	0	1
Family wealth (ref.: poor)					
Relatively poor	1611	0.20	0.40	0	1
Average	1611	0.20	0.40	0	1
Relatively well-off	1611	0.20	0.40	0	1
Well-off	1611	0.20	0.40	0	1
School context					
Daily hours at school	1611	6.74	1.12	2	10
School prevalence of retention	1611	0.05	0.09	0	0.54
School sponsorship fees	1611	4.14	2.98	−4.61	7.74
Non-public school	1611	0.10	0.30	0	1
Lacks library	1611	0.27	0.45	0	1
Lacks science laboratory	1611	0.74	0.44	0	1
Student-teacher ratio: medium	1611	0.34	0.47	0	1
Student-teacher ratio: high	1611	0.34	0.47	0	1

Shen, 2018; Hannum, Kong, & Zhang, 2009; Kong, 2015; Sargent, Kong, & Zhang, 2014; Shen, 2020; Shen, Hu, & Hannum, 2017). The GSCF is a multistage stratified random sample of children between the ages of 9 and 12 in rural Gansu in the year 2000; parents, homeroom teachers, and school principals were also interviewed. Follow-up surveys were conducted in 2004, 2007, 2009, and 2015. We used the data collected in 2000 and 2004, including math and language grades, behavioral problems, and grade retention.

The initial sample size was 2,000 children; 1,918 children were successfully interviewed again in 2004. In our analysis, we included only the participants for which there was information on math, Chinese, and behavioral problems for both surveys. We further restricted our data to cases without missing values at both dependent and independent variables. For the math and language scores, this meant that children had to be enrolled in both years. The sample size for students enrolled in both years with valid score reports is 1,184 for math and 1,181 for language scores. Behavioral problems were measured regardless of enrollment status, and so the analytic sample size for these measures was larger, at 1,611 for both externalizing and internalizing problems. Cases excluded due to missing values were not significantly different from those that were included in the analysis terms of their academic and behavioral outcomes and prevalence of retention. Thus, our estimation is unlikely affected by missing values in the GSCF. We address further the issue of selection due to dropout in robustness checks in Section 4.5.

3.2.1. Measurement

Our dependent variables are educational achievement and behavioral problems. Educational achievement is measured by grades in math and Chinese. As with the CEPS data, grades were reported by teachers based on the children's transcripts. The items for measuring behavioral outcomes in the GSCF were adapted from Achenbach's widely used Child Behavioral Checklist (Achenbach, 1991), which includes *externalizing* and *internalizing behavior problems*[11] (hereafter EBP and IBP). For grade retention information, in the first survey, household heads were asked whether the sample child repeated a grade before 2000 (hereafter *early retention*), and in the second survey between 2000 and 2004 (hereafter *late retention*).

Four sets of independent variables prior to retention are included in the analysis. First, pre-retention *math* and *Chinese grades*, and *externalizing* and *internalizing behavioral problems*, measured in 2000, are included to account for pre-retention differences in educational achievement and behavioral problems. Child characteristics and family background are parallel in the CEPS and the GSCF, and the GSCF also includes measurements from the year 2000 of primary school contexts in which grade retention occurred. To account for school contextual effects, *daily hours at school, school prevalence of grade retention, school type, school sponsorship fees, poor school facilities,* and *student–teacher ratios* are included to predict the likelihood of being retained. *School prevalence of grade retention* is calculated as the total number of repeaters divided by the total number of students. The *school sponsorship fee* is reported by the school principal. Sponsorship fees are set by schools for migrant students without local household registrations, or are "extra" fees parents may pay to help their children to be admitted by reputable schools that have limited admission. School sponsorship fees are log-transformed.[12] Measures of poor school facilities include two variables indicating deprivations: *lacks library* and *lacks science laboratory* are coded 1 to indicated that the facility is lacking and 0 otherwise. The *student–teacher ratio* measures the workload of teachers and available educational resources within a school and reflects the overall teaching and learning environment.

3.2.2. Methods

To analyze the impact of retention on subsequent academic and behavioral outcomes, we utilize longitudinal data to address the selectivity of retention, incorporating pre-retention measurements of demographic characteristics, prior educational achievement, behavioral problems, educational experience, family background, and school contexts. In our analysis, we define *late retention*—retention between the two waves of the survey, 2000 and 2004—as our treatment variable. We consider *early retention*—meaning retention reported in the year 2000, at baseline—as a control variable. We adopt this approach for three reasons. First, this approach allows for more covariates

[11] Externalizing and internalizing behavioral measurements consist of 19 and 17 items, respectively. Each item is reported by children themselves using a 4-point Likert scale, with 1 for "strongly disagree," 2 for "disagree," 3 for "agree," and 4 for "strongly agree." The final scores for both internalizing and externalizing behavioral problems are constructed by summing up each item. Higher scores indicate more severe behavioral problems.

[12] These fees are log-transformed to deal with skewness. Consistent with convention, we add a small constant (0.01) to 0 values such that the log can be taken (see, for example, McDonald, 2015).

to be used to predict late retention while taking into account relevant covariates at baseline. Second, this approach makes it possible to confirm that baseline covariates are temporally prior to treatment. Third, if early retention positively or negatively impacts children's achievement or behavioral problems, retention itself can also be a risk or protective factor for subsequent retention.

For the purposes of this analysis, the first step is to predict the likelihood of being retained between 2000 and 2004 based on these pre-retention covariates measured in 2000. We use binary logistic regression and calculate propensity scores based on this prediction. Because sample sizes depend on the number of valid observations on the dependent variables, we separately estimate the propensity scores by the outcomes of treatment (for more details, please see Appendix B). Second, we conduct kernel-based matching based on propensity scores to balance pre-retention covariates between retained and non-retained students in order to remove retention selectivity based on observed covariates. Kernel-based matching is a nonparametric approach of propensity score matching and yields one-to-many matching, which can be used for calculating the average weight of outcomes for all untreated and treated groups and has been utilized to examine the impacts of grade retention in the United States (Ou & Reynolds, 2010). An advantage of kernel matching is that all untreated cases within a determined span will be utilized; however, results from kernel-based matching may be sensitive to a predetermined bandwidth selection. Different bandwidth and percent of trimming cases are used for testing the robustness of our results.

Under the common support assumption, we compare the outcome variables between treated and untreated groups to obtain the ATT for academic and behavioral outcomes. After matching, we then conduct a balance check to ensure that our matching sufficiently balances the differences between retained and non-retained children based on their observed covariates. We tested balance properties between treated and untreated groups based on Rubin's B (the absolute standardized differences) and R (the variance ratio). The results of our balance check show that Rubin's B and R were all less than 25 percent and between 0.5 and 2 respectively, suggesting a sufficient balance between treated and non-treated groups after matching (Rubin, 2001). We also tested the balance of each variable after matching. Each variable was sufficiently balanced after matching based on the *t*-test. Pre-retention characteristics between retained and non-retained students are similar after matching, thus providing an adequate base for comparing these two groups. We also utilize a difference-in-difference approach, which focuses on the differences in the change of outcomes over a 4-year period between retained and non-retained students (Guo & Fraser, 2014; Heckman, Ichimura, and Todd 1998, 1997). In this way, we eliminate differences in unobserved time-invariant pre-retention characteristics between retained and non-retained students.

A limitation of this study is that our causal inference is based on data from one province. We were precluded from developing causal inferences from the national CEPS data because the sampled children were already in middle school and there is little earlier background information about them. For this reason, it was not possible to control for selectivity of repetition—a lack of information made generating propensity matching scores impossible with the national data. Causal inference based on the GSCF data may not be generalizable at the national scale. However, given literature suggesting that retention tends to be more widely experienced in less-developed areas, Gansu Province is a good illustrative case as a province with a large, rural, impoverished population.

4. Results

4.1. Retention estimates

In Fig. 1, we compile a variety of statistics about grade retention from different sources, including the GSCF and CEPS as well as PISA and official administrative data. We report these sources to give a sense of the variety in estimates and scale of the phenomenon, but it is important to note that they are not directly comparable calculations. Definitions for each estimate are given in the figure.

Fig. 1 underscores the point already discussed: official estimates and school-based reports give a very different picture of the scope of retention than reports from students and families about their own experiences. In the GSCF, the percent of students who repeat a grade according to school principals is around 5.2 percent (total number of repeaters divided by total number of enrolled students in 2000). Yet, 37 percent of students interviewed in the year 2000 reported an experience with repetition. The school reports from Gansu are one-time estimates (period estimates), whereas student or family reports are "lifetime" estimates (cohort estimates), which represent cumulative exposure to

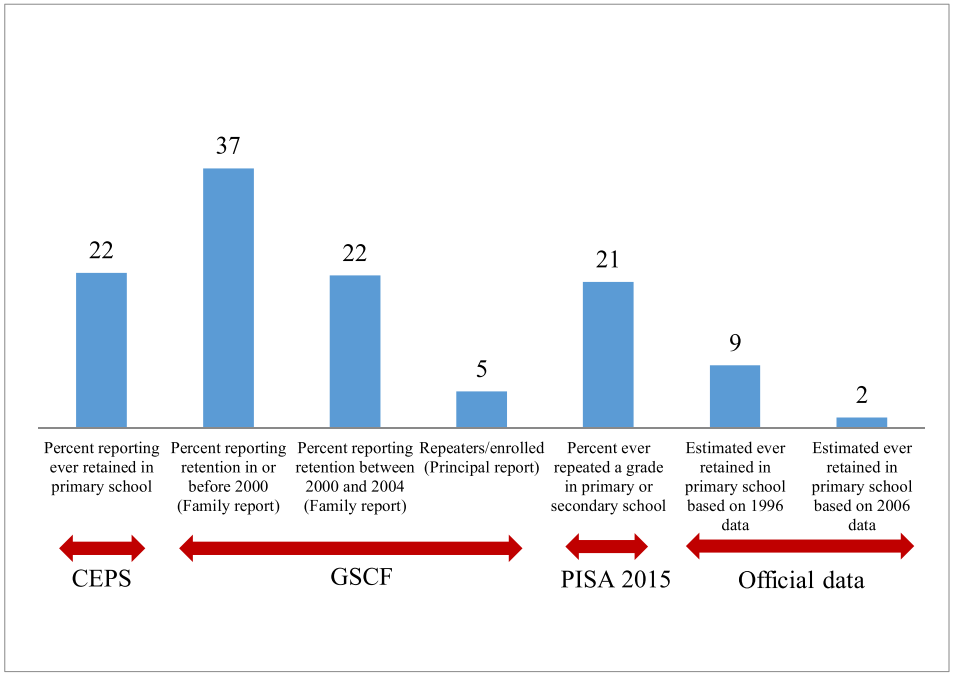


Fig. 1. Grade Retention Statistics from Different Sources.
Note: PISA 2015 includes Beijing, Shanghai, Jiangsu Province, and Guangdong Province. In the CEPS, the question for children reads, "From first grade to sixth grade, how many times have you been retained?" In GSCF, the household head is asked whether the child repeated a grade and which grades the child repeated. For principal-reported prevalence of grade retention in the GSCF, the school principal is asked to provide the number of enrolled male and female students and of male and female repeaters in each grade. The prevalence of grade retention is then calculated as the total number of repeaters divided by the total number of enrolled students. In PISA, 15-year-old students reported whether they repeated a grade in at least one International Standard Classification of Education (ISCED) level. Thus, the prevalence of grade retention in PISA includes both primary and lower secondary education repetition.
Sources: Estimates based on official data are adapted from UNESCO Institute for Statistics (UNESCO Institute for Statistics, 2018). PISA 2015 estimates use OECD data (OECD, 2016).

annual risk of retention over years, so we would expect the lifetime estimates to be larger. In a very different part of China, the percent of students reporting having been retained during primary and lower secondary education is sizable at about 20.8 percent in PISA 2015 (OECD, 2016), which covered some of the wealthiest regions in China (Beijing, Shanghai, Jiangsu Province, and Guangdong Province).

Finally, we calculated a national hypothetical cohort estimate for primary school retention, from grade-specific retention rates. This number represents an estimate of the total retention experience in primary school for a hypothetical cohort experiencing a given year's rate of retention in each grade, with certain assumptions.[13] Fig. 1 shows this number for 2006, which is about the year the CEPS cohort started primary school; it was about 1.5 percent. For the previous year for which data are available, and one perhaps more relevant for national comparison to the Gansu survey, 1996, a corresponding national figure is about 8.9 percent (World Bank, 2020). More recent official estimates are much lower.

In short, official measures are capturing retention occurring in or reported by a given school, for a given year, while student lifetime reports represent their cumulative experience of annual rates, over years and, potentially, across schools. Despite modest and declining official estimates of retention rates, student "lifetime reports" of retention suggest that this phenomenon is significant and worthy of study in China.

Fig. 2 displays the estimated prevalence of grade retention in China according to the CEPS definition of a child reporting having been held back between first and sixth grades. For the CEPS, the data includes information about the locations of middle schools, administrative levels, and regions. We would attain greater precision by measuring the regions and administrative levels of primary schools attended, but this information is not available in the CEPS. Instead, the location of middle school is used as a proxy. Using the CEPS definition, we find that the overall prevalence of grade retention during primary school, despite government strictures on primary school grade retention, is around 22 percent. In the CEPS data, children attending school in city- or county-level central districts, direct-controlled municipalities, provincial capitals, and the eastern region have the lowest prevalence of grade retention, while those enrolled at schools in rural areas and the central part of China have the highest prevalence.[14]

It might be expected that the western region of China should have the highest retention rates, given that this region has long been viewed as the poorest in the country. However, perhaps due in part to poverty alleviation schemes in the west, a study has suggested that government compulsory education expenditures per student in the central region is actually the lowest among western, central, and eastern regions of China (Lei, Qian, & Ma, 2014). This phenomenon is referred to as the "sinking" of the central region of China, and is observed not only in the realm of education, but also in the realm of economic development (Yang & Zhu, 2007). Students at the primary level of education in the central region of China have limited public educational resources compared to western and eastern region of China. It is not surprising, then, that grade retention is more likely to occur in this region.

[13] In technical terms, we created a "synthetic cohort" to convert the period measurement to something more comparable to a lifetime report, with certain assumptions. According to official data, in 2006, which is roughly the year the CEPS cohort began primary school, about 1.22% of first grade student repeated a grade, and 0.10–0.01% for the rest of grade. If a synthetic cohort experienced the same chance of being retained at each grade as mentioned, this synthetic cohort eventually has about 1.5% of students ever being retained, assuming every student can only be retained once and no students transfer in or out. In a conservative estimation, assuming that each grade has 1.22% of students being retained and students can only be retained once and no students transferring in and out, this synthetic cohort eventually has about 7.01% of students ever being retained. Although there is still a discrepancy between official and survey data, the gap is less striking than it might appear at first glance.

[14] Chi-squared tests for associations between grade retention, location of school, administrative level and region are all significant at 0.001 level.

Addressing our first research question, patterns described here are consistent with the expectation that children in less well-off regions of China tend to have higher vulnerability to grade retention. This association is further consistent with the fact that the GSCF (2000) data reveal that 37 percent of rural children aged 9 to 12 in one of China's poorest provinces reported having repeated a grade (see Fig. 1). Overall, these findings indicate that repeating a grade during primary school is a common experience, especially in less resource-rich regions of China.

4.2. Associations of retention with middle school outcomes

Tables 2 and 3 address the descriptive goals of the paper by describing the associations of retention with subsequent outcomes on a national scale. Table 2 displays national estimates of the differences in math grades, language grades, and cognitive test scores for retained and non-retained students, according to CEPS data. Nationally, grade retention is associated with lower educational performance scores (0.391 and 0.358 standard deviations lower), a higher level of depressive symptoms (0.153 points higher on a five-point scale), and lower cognitive test scores (0.507 standard deviations lower). An OLS regression estimation of educational achievement, depressive symptoms, and cognitive test scores with child characteristics, family background, and school dummy variabels using CEPS data is displayed in Table 3. Table 3 shows that for China as a whole, after adjusting for potential confounders, grade retention is significantly associated with lower grades in math, Chinese, and English, with higher levels of depressive symptoms, and with lower cognitive test scores. These national results clearly suggest that grade retention is associated with poorer academic and non-academic outcomes.

4.3. Testing regional differences

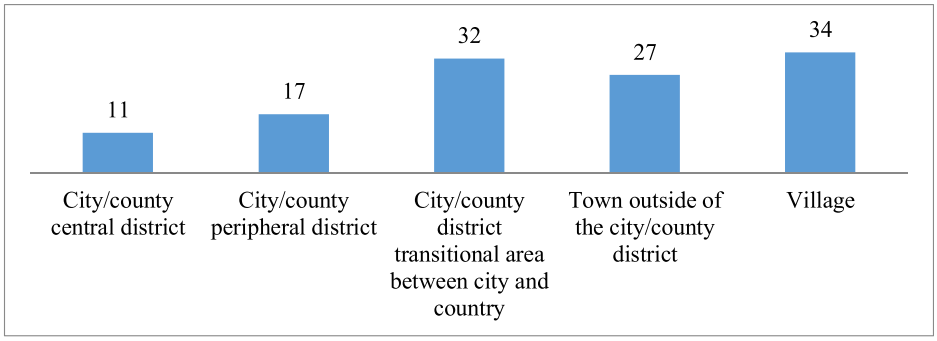
As already noted, Fig. 2 has shown that CEPS students in direct-controlled municipalities (generally wealthier urban settings) report the lowest prevalence of having experienced grade retention, while students in counties (generally poorer rural settings) report the highest prevalence of grade retention. This finding addresses our first research question, are students in less well-off regions of China more vulnerable to grade retention, in the affirmative.

Table 4 shows results addressing our second research question: Do associations of retention with academic and psychosocial outcomes differ in more rural, isolated areas, compared to wealthier urban areas? To address this question, Table 4 shows the same models as Table 3, but adds interaction terms between administrative level and retention to test whether the estimated effects of retention differ depending on the development level of the region, and, by extension, by prevalence.

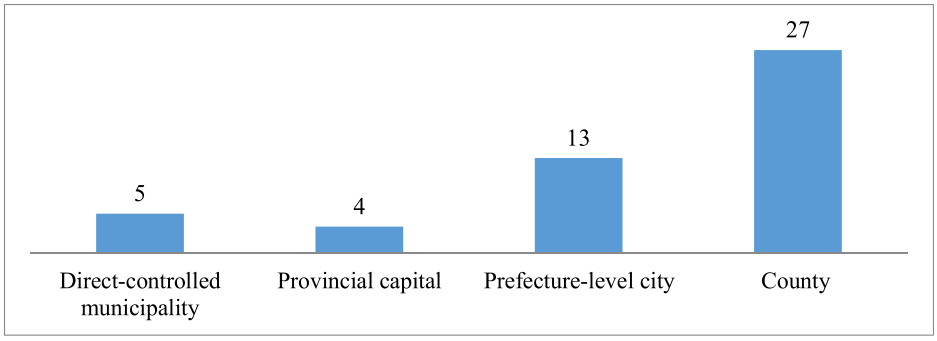
Table 4 offers only suggestive results. Marginally-significant interaction terms indicate that being retained in counties might be associated with more negative implications for some performance measures (Chinese grade, cognitive test scores) than being retained in direct-controlled municipalities. However, there are no differences in math grade and English grade. In contrast, the finding for depressive symptoms shows the opposite pattern, but again is only suggestive: being retained in counties is associated with less of a hit to depressive symptoms than in municipalities, but this finding is only marginally significant. Addressing our second research question, these results do not suggest strong patterns of difference in impact across administrative levels. However, the pattern shown may hint at a tendency for students retained in poorer areas to face less learning support, but also less stigma.

4.4. Impact of grade retention on subsequent outcomes

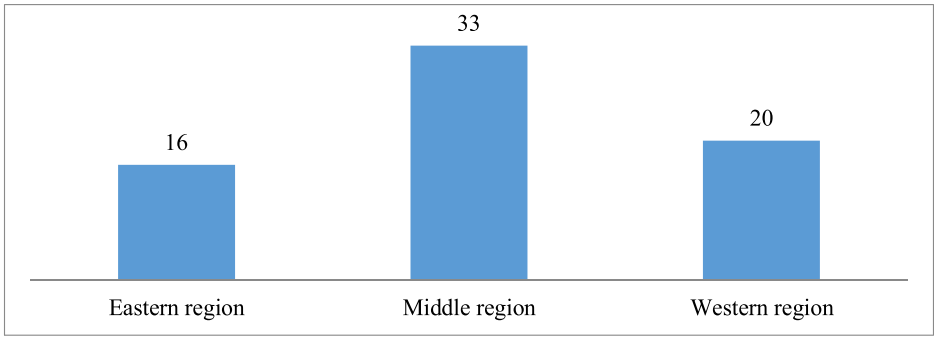
In Table 5, we address our third research question: is there an estimable causal impact of retention on academic and behavioral outcomes? Table 5 presents the estimated impact of grade retention on educational achievement and behavioral problems using kernel-based matching and



Panel (a): Percent reporting ever been retained in primary school by school location



Panel (b): Percent reporting ever been retained in primary school by administrative level



Panel (c): Percent reporting ever been retained in primary school by region.

Fig. 2. Percent Reporting Ever Been Retained in Primary School by Location of Middle School.
Note: School location, administrative level, and region are categories used in the CEPS. *Source*: CEPS.

Table 2
Mean Differences in Educational Achievement, Depression, and Cognitive Test Scores between Non-Retained and Retained Students in the CEPS.

	Non-retained	Retained	Difference	P-values
Math grade	0.065	−0.326	0.391	0.000***
Chinese grade	0.061	−0.297	0.358	0.000***
English grade	0.079	−0.398	0.477	0.000***
Depressive symptoms	2.065	2.218	−0.153	0.000***
Cognitive test	0.084	−0.23	0.507	0.000***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$.

difference-in-difference analysis. The unadjusted mean differences in educational achievement and behavioral problems in 2000 show substantial differences between retained and non-retained students *before* grade retention. Retained students, on average, had 0.431 more externalizing behavioral problems (0.431 units) than non-retained students in 2000. Retained students also had greater internalizing behavioral problems (0.357 units) and lower math and Chinese grades (−0.622 and −0.586 standardized deviations) than all non-treated groups.

In 2004, the externalizing behavioral problems of retained students persisted after grade retention, but the gap between these students and others narrowed significantly, and improvements for all outcomes were better for retained than for other students, according to the unmatched difference-in-difference estimator. The difference in externalizing behavioral problems between retained and non-retained students decreased from 0.431 before grade retention to 0.152 after. The gap between retained and non-retained students in internalizing behavioral problems and educational achievement disappeared. As for educational achievement, math and Chinese grades improved after grade retention and retained students, on average, gained 0.523 and 0.492 points more in math and Chinese grades, respectively, compared to non-retained students. Thus, the unadjusted mean differences in educational achievement and behavioral problems before and after grade retention reveal that students can improve with retention in both academic and non-academic outcomes.

Table 3
OLS Estimation of Educational Achievement, Depression, and Cognitive Test Scores with Grade Retention, CEPS.

	(1) Math	(2) Chinese	(3) English	(4) Depressive symptoms	(5) Cognitive test
Retention (ref.: not retained)	−0.093***	−0.074***	−0.141***	0.060**	−0.092***
	(0.021)	(0.019)	(0.020)	(0.020)	(0.018)
Gender					
Girl	0.056***	0.420***	0.432***	0.065***	−0.041***
	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)
Missing	−0.326***	−0.103*	−0.141**	0.125*	−0.121*
	(0.056)	(0.052)	(0.053)	(0.053)	(0.048)
Only child	0.019	0.011	0.030*	−0.006	−0.021
	(0.016)	(0.015)	(0.015)	(0.015)	(0.014)
School transfer	0.067***	0.044**	0.037**	0.051***	0.047***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)
Age upon entering primary school	0.038***	0.034***	0.039***	0.009	0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
Non-kindergarten attendance					
No	−0.070***	−0.051***	−0.037*	0.053***	−0.092***
	(0.016)	(0.015)	(0.016)	(0.016)	(0.014)
Missing	−0.211*	−0.184*	−0.350***	0.027	−0.223**
	(0.097)	(0.090)	(0.093)	(0.093)	(0.084)
Parents' educational expectations (ref.: less than college)					
College and above	0.606***	0.462***	0.560***	−0.079***	0.318***
	(0.015)	(0.014)	(0.015)	(0.015)	(0.013)
Missing	0.319***	0.197***	0.249***	−0.027	0.174***
	(0.054)	(0.050)	(0.052)	(0.051)	(0.047)
Constant	0.771***	0.156	0.205	2.153***	1.244***
	(0.162)	(0.151)	(0.155)	(0.154)	(0.141)
Observations	18,143	18,148	18,150	18,275	18,604
R-squared	0.301	0.393	0.362	0.067	0.279

Note: Standard errors are reported in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Control variables included in the model are ethnicity, age of children, child migration and hukou status, health problems, parental migration status, parents' educational expectations, father's and mother's levels of education, family wealth, grade, semester, and school dummies.

Table 4
OLS Estimation of Educational Achievement, Depression, and Cognitive Test Scores with Interaction Terms between Administrative Level and Grade Retention, CEPS.

	(1) Math	(2) Chinese	(3) English	(4) Depressive symptoms	(5) Cognitive test
Retention (ref: not retained)	−0.061	0.032	−0.141 +	0.168*	0.013
	(0.076)	(0.071)	(0.073)	(0.071)	(0.065)
Retention#provincial capitals	0.095	−0.179 +	0.034	0.037	−0.115
	(0.104)	(0.097)	(0.100)	(0.098)	(0.089)
Retention#prefecture-level city	0.090	−0.015	0.075	−0.137	−0.088
	(0.092)	(0.086)	(0.089)	(0.086)	(0.079)
Retention#county	−0.067	−0.123 +	−0.016	−0.125 +	−0.117 +
	(0.079)	(0.073)	(0.076)	(0.073)	(0.067)
Constant	0.782***	0.155	0.207	2.154***	1.245***
	(0.161)	(0.150)	(0.155)	(0.153)	(0.140)
Observations	18,143	18,148	18,150	18,275	18,604
R-squared	0.301	0.394	0.362	0.067	0.279

Note: Standard errors are reported in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. Other variables in the model are the same as control variables listed in the table note for Table 3. Direct-controlled municipality is the reference category for region.

However, the results are substantially different when we adjust for differences by sample matching and estimate the average treatment effects for the treated (ATT). We use kernel matching to adjust for pre-existing differences between retained and non-retained students, and successfully matched the sample such that pre-retention differences in educational achievement and behavioral problems narrowed. For instance, the unadjusted mean difference in pre-retention externalizing behavioral problems between retained and non-retained students is 0.431, but is reduced to −0.029 after matching. Using this design, the ATT on externalizing problems in 2004 is not statistically significant after matching. The average treatment effects for the change in externalizing behavioral problems between 2000 and 2004 show a similar pattern. With regard to difference-in-difference estimators, the unadjusted mean difference in the change of externalizing behavioral problems between retained and non-retained students is around −0.279 units and is significant at the $p < 0.001$ level, but the treatment effect for the treated is around 0.137 units and is no longer statistically significant. There is a similar pattern for internalizing behavioral problems and math and Chinese grades. In other words, after retention selectivity is taken into account, the effects of grade retention erode.[15] Overall, after balancing the pre-retention observed and unobserved characteristics between retained and non-retained students, the impacts of grade retention on both academic and non-academic outcomes are negligible.

[15] We also conducted sensitive analysis with the same model specification, but with different bandwidth and trimming strategies (Guo & Fraser, 2014). Our analysis suggests that the ATT changes slightly with a different bandwidth (0. 01, 01, 0.5) and trimming strategies (5%, 10% and 15%); nevertheless, the overall pattern shows no significant impact of grade retention on both educational achievement and behavioral problems. The only exception is the impact of grade retention on externalizing behavioral problems, as the impact becomes more pronounced when a higher percentage of cases is trimmed.

Table 5
Estimations of the ATT for Educational Achievement and Behavioral Problems using Kernel-Based Matching and Difference-In-Difference Estimator.

Variable	Sample	Treated	Controls	Difference	SE	T-stat
EBP[a], 2004	Unmatched[c]	0.111	−0.041	0.152	0.061	2.49*
	ATT[d]	0.120	0.013	0.107	0.087	1.23
EBP, 2000	Unmatched	0.344	−0.087	0.431	0.061	7.11***
	ATT	0.316	0.345	−0.029	0.088	−0.33
IBP[b], 2004	Unmatched	0.013	−0.004	0.017	0.061	0.28
	ATT	0.022	0.026	−0.004	0.078	−0.05
IBP, 2000	Unmatched	0.284	−0.073	0.357	0.060	5.91***
	ATT	0.253	0.270	−0.018	0.089	−0.20
Math grade, 2004	Unmatched	−0.080	0.019	−0.099	0.064	−1.54
	ATT	−0.071	−0.057	−0.015	0.084	−0.18
Math grade, 2000	Unmatched	−0.432	0.191	−0.622	0.062	−10.10***
	ATT	−0.353	−0.415	0.062	0.078	0.79
Chinese grade, 2004	Unmatched	−0.077	0.018	−0.094	0.064	−1.47
	ATT	−0.056	−0.132	0.076	0.087	0.87
Chinese grade, 2000	Unmatched	−0.403	0.183	−0.586	0.064	−9.21***
	ATT	−0.321	−0.374	0.053	0.090	0.59
DID estimator						
EBP	Unmatched	−0.233	0.046	−0.279	0.083	−3.38***
	ATT	−0.195	−0.332	0.137	0.121	1.13
IBP	Unmatched	−0.271	0.069	−0.340	0.085	−4.00***
	ATT	−0.230	−0.244	0.014	0.119	0.12
Math grade	Unmatched	0.352	−0.172	0.523	0.082	6.37***
	ATT	0.282	0.358	−0.077	0.110	−0.70
Chinese grade	Unmatched	0.326	−0.166	0.492	0.082	5.99***
	ATT	0.265	0.242	0.023	0.116	0.20

Note: Standard errors for ATT's are estimated via bootstrapping. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$.
[a] EBP refers to externalizing behavioral problems.
[b] IBP refers to internalizing behavioral problems.
[c] Unmatched indicates unadjusted mean differences.
[d] ATT refers to average treatment effects for the treated.

4.5. Robustness checks: school transfer and dropout

There are two issues, school transfer and school dropout, that might undermine our interpretations. School transfer is particularly important in China due to rising family migration and school consolidation in rural communities (Hannum, Liu, & Wang, 2018; Kipnis, 2006). Students moving to better-resourced schools after consolidation or migration may be particularly likely to repeat a grade and may be vulnerable in other unmeasured ways. There is no school transfer information in the GSCF. But in the CEPS, school transfer is common: about 32 percent of participating students reported that they transferred during primary school. Among those who did not transfer schools, 14.5 percent reported having repeated a grade, while this figure for those who experienced a school transfer is 38.5 percent. Table 6 displays the same model specification as Table 3, but provides estimates separately by school transfer status. Results show that retained students, on average, display lower educational achievement, lower test scores, and more depressive symptoms than non-retained students regardless of school transfer status. It is important to comment on the stark differences in retention across school transfer status beyond their role in a robustness check. These differences suggest that transfer may be one significant reason that retention or repetition goes unreported in official statistics, as students enter new schools for the first time.

School dropout is another factor that may bias our estimation, as a portion of the students in the GSCF were not enrolled in school in 2004. Information on achievement is not available for those who dropped out in 2004, but information on behavioral problems was collected regardless of enrollment status. Kernel-based matching and difference-in-difference with the same model specification as Table 5 are conducted by enrollment status in 2004 to examine whether the impact of grade retention on behavioral problems varies by enrollment status. Table 7 shows the impact of grade retention on externalizing and internalizing behavioral problems by enrollment status in 2004. For the unmatched sample, we see that there is no obvious impact of grade retention on behavioral problems among those who dropped out in 2004, while we see a strong relationship between grade retention and behavioral problems among those who stayed in school. Overall, retained students who stayed in school in 2004 exhibited a decrease in externalizing behavioral problems between 2000 and 2004. However, the ATT for grade retention shows an estimated causal impact increasing externalizing behavioral problems by 0.224 units ($p < 0.05$), but no significant impact on internalizing behavioral problems. The matched result suggests that retained students who stayed in school exhibited more externalizing behavioral problems, while those who dropped out did not.

We also conducted dropout analysis by predicting the likelihood of dropout in 2004 based on retention between 2000 and 2004 and pre-retention covariates measured in 2000 using binary logistic regression (results available upon request). Results suggest that children with low

Table 6
Coefficients of Grade Retention on Academic and Non-Academic Outcomes by School Transfer Experience, CEPS.

	No school transfer	School transfer
Math	−0.133***	−0.066*
	(0.030)	(0.030)
Chinese	−0.075**	−0.080**
	(0.028)	(0.028)
English	−0.156***	−0.135***
	(0.028)	(0.029)
Depression	0.080**	0.040
	(0.028)	(0.026)
Cognitive test	−0.107***	−0.085***
	(0.026)	(0.024)

Note: Standard errors are reported in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. All control variables listed in the table note for Table 3 are included in the model.

Table 7
Kernel-Based Matching and Difference-In-Difference Estimators by Enrollment Status in 2004, GSCF. [a] and [b]

		Enrolled		Non-enrolled	
	Sample	Difference	T-stat	Difference	T-stat
EBP, 2004	Unmatched	0.171	2.71***	0.134	0.56
	ATT	0.156	1.73*	−0.012	−0.03
EBP, 2000	Unmatched	0.434	6.83***	0.328	1.48
	ATT	−0.067	−0.70	0.000	0.00
IBP, 2004	Unmatched	0.045	0.71	−0.108	−0.46
	ATT	0.054	0.67	−0.107	−0.32
IBP, 2000	Unmatched	0.359	5.69***	0.321	1.43
	ATT	−0.033	−0.36	−0.064	−0.21
DID estimators					
EBP	Unmatched	−0.263	−3.06**	−0.194	−0.62
	ATT	0.224	1.70*	−0.012	−0.03
IBP	Unmatched	−0.315	−3.54***	−0.429	−1.42
	ATT	0.087	0.068	−0.042	−0.08

Note: Standard errors for ATT's are estimated via bootstrapping. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$.
[a] Unmatched indicates unadjusted mean differences.
[b] ATT refers to average treatment effects for the treated.

educational expectations from mothers, with poorly-educated parents, who attend schools without a library, and who have a high prevalence of grade retention are more likely to drop out. Interestingly, children who experienced late retention were more likely to still be in school in 2004. This implies that instead of retention being a precursor to quitting school, children or their families may consider being retained as a chance to catch up, or as a means to attain a certain minimum level of education. Taken together, our estimation of the impacts of grade retention on educational achievement shows that although children from advantaged families and children who attend schools with better facilities are more likely to stay in the education system, the benefits of grade retention on educational achievement are trivial.

In sum, in the CEPS analysis, grade retention during primary school is associated with poorer outcomes for every domain of wellbeing after controlling for demographic characteristics, educational experience, family background, and school context. Children in less developed regions are more vulnerable to retention, but evidence about whether the experience of retention carries different implications in the poorest regions is only suggestive. In the GSCF, there are improvements in both academic and behavioral outcomes with retention in the unadjusted mean differences, but these improvements disappear after adjusting for differences in observed and unobserved characteristics prior to retention between retained and non-retained students. At the same time, school dropout has a slight influence on our estimation of the impact of grade retention on externalizing behavioral problems, suggesting that students who remain in school after grade retention exhibit more behavioral problems.

5. Conclusion and discussion

This paper has investigated the prevalence and correlates of grade retention and its impact on children's academic and non-academic outcomes with nationally representative, cross-sectional data and with longitudinal data from one of the poorest provinces in China. Despite very low reports of officially recognized retention in China, our results show that a non-trivial minority of children experience grade retention during elementary school. Children in less-developed regions are more likely to do so: the prevalence of grade retention is higher in rural regions, suggesting that retention is inversely associated with economic development and educational resources. This finding is consistent with earlier work in a variety of national contexts that indicates a higher prevalence of retention in rural and remote areas (Brophy, 2006). In contexts of high resource deprivation, retention may be one of few available options to support children who are struggling.

Moreover, the discrepancy between the prevalence of retention that is officially captured and what is reported by children and family is striking. We speculate that the high prevalence of child- and family-reported grade retention may be partially linked to rural school mergers, family migration, and the availability of school choice in China. Official data relies on a static measurement of grade retention at a given school and may not accurately reflect the prevalence of grade retention if there is a vast flow of school transfer that is not well documented. Children changing schools may repeat a grade and this repetition may go unrecorded in the receiving schools' records. If this is the case, both the school merger program launched by the government and school transfers initiated by migrating parents may be contributing significantly to grade retention that is unrecorded in official statistics. This explanation is consistent with the finding in our robustness checks that grade retention is much more frequently reported in the CEPS data among children who transferred schools—38.5 percent of whom repeated—than among other children—14.5 percent of whom repeated.

The disconnect in scale of the phenomenon implied by official estimates and student- or family-reported estimates is important to consider beyond the Chinese context, because official statistics are the primary data source for describing the scale of grade retention worldwide. Many developing countries are undergoing vast flows of migration and rapid changes in educational infrastructure, which may lead to inaccurate and partial official representations of grade retention. As a result, grade retention issues may be neglected in educational policy discussions and may not attract the attention, resources, and services that are necessary for helping retained children.

Finally, as seen in other cases, retained children in China fare more poorly than non-retained children in subsequent educational performance and behavioral problems, and our national analysis shows that this association with poorer outcomes is not definitively different among students in poorer and wealthier regions. It is important to note, however, that the majority of our findings are consistent with retention being a "red flag" indicating risk of educational problems and behavioral problems, rather than a cause of these problems.[16] Our longitudinal analysis with unmatched samples suggests that retained children *can* gain ground in academic and behavioral domains. Our causal analyses indicate that the experience of grade retention itself, for the most part, has neither harmful nor beneficial impacts when we compare similar samples. These results suggest that retention, quite possibly the only strategy available to poor parents and low-resource schools in China, has limited potential as an academic or behavioral intervention. Collectively, both the main findings and suggestive evidence linking retention to certain behavioral problems suggest that additional counselling services may be an appropriate strategy for supporting retained children.

CRediT authorship contribution statement

Li-Chung Hu: Conceptualization, Methodology, Formal analysis, Writing - original draft. **Emily Hannum:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[16] One possible exception to this statement emerged in one robustness check: there is suggestive evidence that retention is linked to more externalizing behavioral problems among retained students who did not drop out.

Acknowledgements

An earlier version of this paper was presented at the International Sociological Association 2016 RC28 (Social Stratification and Mobility) Meeting in Singapore. The authors gratefully acknowledge the comments and suggestions given by Lingxin Hao and Tony Tam on earlier drafts. The author's travel expenses were funded by the Ministry of Science and Technology, Taiwan (MOST 105-2914-I-004-015-A1).

The authors gratefully acknowledge the Hong Kong Research Grants Council (HSSPFS 36600014), the United Kingdom Economic and Social Research Council and Department for International Development (ESRC RES-167-25-0250), the Spencer Foundation Small and Major Grants Programs, the World Bank, and NIH Grants 1R01TW005930-01 and 5R01TW005930-02 for supporting the data collection and processing for the Gansu Survey of Children and Families. The authors acknowledge additional research support from the Population Studies Center at the University of Pennsylvania (National Institutes of Health - National Institute on Aging, Grant number P30 AG12836, the Boettner Center for Pensions and Retirement Security at the University of Pennsylvania, and National Institutes of Health – National Institute of Child Health and Development Population Research Infrastructure Program R24 HD-044964). This research also uses data from the China Education Panel Surveys, collected and distributed by the National Survey Research Center (NSRC) at Renmin University of China. We thank the Renmin University Scientific Research Foundation, the Social Survey Foundation of China's National Survey Research Center, and the United States National Science Foundation (Award 1259530) for supporting the China Education Panel Survey.

Appendix A. Definition of Measures Used in the CEPS and GSCF

	CEPS	GSCF
Dependent variables		
Math grade	2013 transcript; grade in the last semester	2004 transcript; grade in the last semester
Language grade	2013 transcript; grade in the last semester	2004 transcript; grade in the last semester
English grade	2013 transcript; grade in the last semester	n/a
Behavioral measurements	Depressive symptoms	2004 Achenbach's Child Behavioral Checklist
Cognitive test	Designed by researchers; standardized tests; curriculum-irrelevant	n/a
Independent variables		
Pre-retained grade		
Math grade	n/a	2000 transcript; grade in the last semester
Language grade	n/a	2000- transcript; grade in the last semester
Pre-retained behavioral measurements		
Behavioral measurements	n/a	2000 Achenbach's Child Behavioral Checklist
Child characteristics		
Girl	Reported by children; 0 = boy; 1 = girl; 2 = missing	Reported by household head; 0 = boy; 1 = girl
Only child	0 = no; 1 = yes	0 = no; 1 = yes
Han	0 = Non-Han; 1 = Han	n/a
Age	11–17 years old in 2013–2014	9–12 years old in 2000
Child migration status	0 = local non-immigrant; 1 = intra-provincial; 2 = inter-provincial	n/a
Child current hukou type	0 = non-agricultural; 1 = agricultural	n/a
Health problems	Reported by parents; severe illness before attending primary school; 0 = no; 1 = yes; 2 = missing	Reported by household head; any chronic diseases or disable; 0 = no; 1 = yes
Grade	0 = 7^{th} grade; 1 = 9^{th} grade	n/a
Semester	0 = spring semester, 2014;1 = fall semester, 2013	n/a
Educational experience		
Grade retention	Reported by children; 0 = non-retained; 1 = retained	Reported by household head; 0 = non-retained; 1 = retained
School transfer	0 = no; 1 = yes	n/a
Age upon entering primary school	Reported by children	Reported by household head
Non-kindergarten attendance	0 = no; 1 = yes; 2 = missing	0 = no; 1 = yes
Family background		
Parental migration status	0 = both parents at home; 1 = only mother at home; 2 = only father at home; 3 = neither parent at home	0 = parents are always at home; 1 = other
Parents' educational expectations	Mother's or father's educational expectation; 0 = less than college; 1 = college and above; 2 = missing	Mother's educational expectation. Reported by mothers
Father's and mother's education	7 categories: 0 = no schooling; elementary school; 1 = middle school; 2 = secondary specialized school/technical school; 3 = vocational high school; 4 = high school; 5 = junior college; 6 = college and above school or above	3 categories: 0 = less than primary; 1 = above primary and less than middle school; 2 = middle school and above
Family wealth	Parental self-perceived wealth before child enrolled in elementary school; 0 (poorest)-3 (relatively well-off/well-off); 4 = missing	Aggregate market value of agricultural products, fixed assets, and durable goods; reported by household head; 0 (poorest)-4 (well-off)
School context		
Administrative level	4 categories: 0 = direct-controlled municipality; 1 = provincial capital; 2 = prefecture-level city; 3 = county	n/a
Daily hours at school	n/a	Daily hours at school
School sponsorship fees	n/a	Total amount of money (logged)
School type	n/a	0 = public; 1 = non-public
Lacks library	n/a	0 = yes; 1 = no
Lacks science laboratory	n/a	0 = yes; 1 = no
School prevalence of grade retention	n/a	Number of repeaters divided by number of students
Student-teacher ratio	n/a	Number of students divided by number of teachers, then created three dummy variables, low, medium, and high based on three quintiles of student–teacher ratio.

Note: Math, Chinese, and English grades are standardized by schools in both CEPS and GSCF. School contextual variables are all reported by school principals in GSCF.

Appendix B. Binary Logistic Estimation for the Propensity of Being Retained between 2000 and 2004 on Pre-retained Grade and Behavioral Problems, Child Characteristics, Family Background, and School Contexts in 2000 Using the GSCF by the Outcomes of Treatment

	(1)	(2)	(3)
Outcome of treatment	EBP and IBP	Math grade	Chinese grade
Girl	1.004	0.822	0.834
	(0.142)	(0.136)	(0.138)
Child age	0.698***	0.737***	0.734***
	(0.054)	(0.067)	(0.067)
Only child	0.583	0.654	0.648
	(0.205)	(0.263)	(0.261)
Retention, 2000	1.401*	1.315	1.320
	(0.222)	(0.242)	(0.243)
Age upon entering primary school	1.162+	1.201*	1.204*
	(0.090)	(0.109)	(0.110)
Non-kindergarten attendance	1.267	1.125	1.165
	(0.222)	(0.231)	(0.241)
Health problems	1.611	2.109+	2.206*
	(0.522)	(0.832)	(0.878)
Math grade, 2000	0.722**	0.584***	0.586***
	(0.079)	(0.076)	(0.076)
Chinese grade, 2000	0.720**	0.747*	0.742*
	(0.080)	(0.095)	(0.094)
EBP, 2000	1.284+	1.362*	1.341*
	(0.167)	(0.204)	(0.200)
IBP, 2000	0.918	0.854	0.865
	(0.120)	(0.129)	(0.131)
Father or mother are away from home	1.012	1.142	1.139
	(0.250)	(0.326)	(0.325)
Mother's educational expectation	0.952+	0.931*	0.934*
	(0.024)	(0.028)	(0.028)
Father's education (ref.: less than primary school)			
Above primary school and below middle school	0.527**	0.536*	0.531*
	(0.118)	(0.136)	(0.135)
Middle school and above	0.766+	0.781	0.780
	(0.122)	(0.146)	(0.146)
Mother's education (ref.: less than primary school)			
Above primary school and below middle school	0.508*	0.513*	0.554+
	(0.140)	(0.162)	(0.172)
Middle school and above	0.577*	0.530*	0.535*
	(0.125)	(0.133)	(0.134)
Wealth quintile (ref.: poor)			
Relatively poor	0.777	0.702	0.718
	(0.163)	(0.175)	(0.179)
Average	1.104	1.070	1.120
	(0.236)	(0.274)	(0.287)
Relatively well-off	1.190	1.033	1.048
	(0.267)	(0.276)	(0.282)
Well-off	0.965	0.985	0.998
	(0.232)	(0.271)	(0.276)
Daily hours at school	1.144*	1.159+	1.160+
	(0.072)	(0.089)	(0.089)
School sponsorship	0.949*	0.941*	0.941*
	(0.024)	(0.027)	(0.027)
Non-public school	0.479**	0.365***	0.366***
	(0.117)	(0.108)	(0.108)
Lacks library	3.788***	4.351***	4.460***
	(0.665)	(0.925)	(0.948)
Lacks science laboratory	0.816	0.729	0.741
	(0.156)	(0.161)	(0.164)
School prevalence of retention	5.908*	4.548	4.509
	(5.192)	(4.667)	(4.619)
Student-teacher ratio (ref.: low)			
Medium	1.485*	1.595*	1.588*
	(0.292)	(0.372)	(0.369)
High	1.445*	1.357	1.338
	(0.267)	(0.295)	(0.291)
Constant	0.664	0.513	0.447
	(0.698)	(0.617)	(0.538)
LR chi2 (29)	348.93	320.08	317.28
Prob > chi2	0.000	0.000	0.000
Pseudo R2	0.207	0.246	0.244
Observations	1,611	1,184	1,181

Note: EBP refers to externalizing behavioral problems and IBP refers to internalizing behavioral problems. Standard errors are reported in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$.

References

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile.* VT: Department of Psychiatry, University of Vermont Burlington.

Alderman, H., Behrman, J. R., Lavy, V., & Menon, R. (2001). Child health and school enrollment: A longitudinal analysis. *Journal of Human Resources, 36*(1), 185–205.

Alexander, K., Entwisle, D., & Kabbani, N. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record, 103*(5), 760–822.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis, 31*(4), 480–499.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology, 21*(1), 1–20.

Andrew, M. (2014). The scarring effects of primary-grade retention? A study of cumulative advantage in the educational career. *Social Forces, 93*(2), 653–685.

Audas, R., & Willms, J. D. (2002). *Engagement and dropping out of school: A life-course perspective.* Hull, Quebec: Human Resources Development Canada. http://sbisrvntweb.uqac.ca/archivage/15292281.pdf.

Bali, V. A., Anagnostopoulos, D., & Roberts, R. (2005). Toward a political explanation of grade retention. *Educational Evaluation and Policy Analysis, 27*(2), 133–155.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics, 121*(4), 1437–1472.

Brophy, J. (2006). *Grade repetition. Education Policy Series 6*. Paris: International Academy of Education and International Institute for Educational Planning.

Chen, X., Liu, C., Zhang, L., Shi, Y., & Rozelle, S. (2010). Does taking one step back get you two steps forward? Grade retention and school performance in poor areas in rural China. *International Journal of Educational Development, 30*(6), 544–559.

Cherng, H.-Y. S., & Hannum, E. (2013). Community poverty, industrialization, and educational gender gaps in rural China. *Social forces, 92*(2), 659–690.

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement, 24*(4), 417–434.

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*(3), 323–347.

Guèvremont, A., Roos, N. P., & Brownell, M. (2007). Predictors and consequences of grade retention: Examining data from Manitoba, Canada. *Canadian Journal of School Psychology, 22*(1), 50–67.

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications: Statistical methods and applications.* Los Angeles: Sage Publications.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies, 65*(2), 261–294.

Hannum, E. C. (1999). Political change and the urban-rural gap in basic education in China, 1949–1990. *Comparative education review, 43*(2), 193–211.

Hannum, E. C., & Hu, L.-C. (2017). Chronic undernutrition, short-term hunger, and student functioning in rural northwest China. *International Journal of Educational Development, 54*, 26–38.

Hannum, E. C., Hu, L.-C., & Shen, W. (2018). *Short- and long-term outcomes of the left behind in China: Education, well-being and life opportunities." Background paper prepared for the 2019 Global Education Monitoring Report ED/GEMR/MRT/2018/P1/2. Programme and Meeting Document.* Paris: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000266050.

Hannum, E. C., Kong, P. A., & Zhang, Y. (2009). Family sources of educational gender inequality in rural China: A critical assessment. *International journal of educational development, 29*(5), 474–486.

Hannum, E. C., Liu, X., & Wang, F. (2018). Estimating the effects of educational system consolidation: The case of China's rural school closure initiative. *SSRN Working Paper, April 13*. https://doi.org/10.2139/ssrn.3140132https://ssrn.com/abstract=3140132, Accessed date: 28 March 2020.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies, 64*(4), 605–654.

Holmes, C. T., et al. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard, & M. L. Smith (Eds.). *Flunking grades: Research and policies on retention* (pp. 16–33). London: Falmer Press.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205–224.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy. *Journal of the American Statistical Association, 101*(475), 901–910.

Hu, L. -C., & Hannum , E. (in press). Correlates and implications of grade retention in rural Northwest China. In P. Kong, G. A. Postiglione, & E. Hannum (Eds.), China's rural education in transition. Routledge Press.

Hungi, N. (2010). *What are the levels and trends in grade repetition? SACMEQ Policy Issues Series.* Paris: UNESCO/IIEP. http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/policy-issue-series/005-sacmeqpolicyissuesseries-graderepetition.pdf.

Ikeda, M., García, E., et al. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies, 2013*(1), 1–47.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244.

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1*(3), 33–58.

Jamison, D. T. (1986). Child malnutrition and school performance in China. *Journal of Development Economics, 20*(2), 299–309.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30*(3), 420–437.

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools, 39*(4), 441–457.

Kipnis, A. (2006). School consolidation in rural China. *Development Bulletin, 70*, 123–125.

Kong, P. A. (2015). *Parenting, education, and social mobility in rural China: Cultivating dragons and phoenixes.* London and New York: Routledge.

Lei, W., Qian, J., & Ma, H. (2014). Studies on compulsory education investment sinking in central China. *Education & Economy, 6*, 1–7.

Liu, C., Zhang, L., Luo, R., Rozelle, S., & Loyalka, P. (2010). The effect of primary school mergers on academic performance of students in rural China. *International Journal of Educational Development, 30*(6), 570–585.

Lorence, J. (2006). Retention and academic achievement research revisited from a United States perspective. *International Education Journal, 7*(5), 731–777.

Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology, 101*(1), 101–114.

McDonald, J. H. (2015). Data transformations. *Handbook of biological statistics* (pp. 140–144). Baltimore: Sparky House Publishing.

McEwan, P. J., & Shapiro, J. S. (2008). The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. *Journal of Human Resources, 43*(1), 1–29.

Mei, H., Jiang, Q., Xiang, Y., & Song, X. (2015). School consolidation: Whither China's rural education? *Asian Social Work and Policy Review, 9*(2), 138–150.

Nagin, D. S., Pagani, L., Tremblay, R. E., & Vitaro, F. (2003). Life course turning points: The effect of grade retention on physical aggression. *Development and Psychopathology, 15*(02), 343–361.

National Survey Research Center (2020). Introduction to the China Education Panel Survey. *China Education Panel Survey*Beijing: Renmin University of Chinahttps://ceps.ruc.edu.cn/index.php?r=index/introduce, Accessed date: 27 March 2020.

OECD (2011). *When students repeat grades or are transferred out of school: What does it mean for education systems? PISA in Focus6*PARIS: OECD Publishing. https://doi.org/10.1787/5k9h362n5z45-en.

OECD (2016). Annex B1.6 Results (tables): Socio-economic status, student performance and students'attitudes towards science. *PISA 2015 results, Volume I: Excellence and equity in education*. https://doi.org/10.1787/9789264266490-enhttp://statlinks.oecdcode.org/982016061p1t008.xlsx.

Ou, S.-R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis, 32*(1), 118–139.

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology, 13*(2), 297–315.

Qin, H. (2008). School choice in China. *Frontiers of Education in China, 3*(3), 331–345.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3–4), 169–188.

Ryff, C. D. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science, 4*(4), 99–104.

Sargent, T., Kong, P., & Zhang, Y. (2014). Home environment and educational transitions on the path to college in rural northwest China. *International Journal of Educational Development, 34*, 98–109.

Shen, W. (2020). A Tangled Web: The reciprocal relationship between depression and educational outcomes in China. *Social Science Research, 85*, 102353, 1–16.

Shen, W., Hu, L.-C., & Hannum, E. (2017). Cumulative adversity, childhood behavioral problems, and educational mobility in China's poorest rural communities. *Chinese Journal of Sociology, 34*, 491–517.

Stearns, E., Moller, S., Blau, J., & Potochnick, S. (2007). Staying back and dropping out: The relationship between grade retention and school dropout. *Sociology of Education, 80*(3), 210–240.

UNESCO Institute for Statistics. (2009). *Education indicators: Technical guidelines*. Montreal: UNESCO-UIS Publishing. http://uis.unesco.org/sites/default/files/documents/education-indicators-technical-guidelines-en_0.pdf.

UNESCO Institute for Statistics. (2012). *Opportunities lost: The impact of grade repetition and early school leaving. Global Education Digest*. Montreal: UNESCO-UIS Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000218449.

UNESCO Institute for Statistics. (2018). UIS Database, http://data.uis.unesco.org, [May 11, 2018].

Tsang, M. C. (2003). School choice in the People's Republic of China. In D. N. Plank, & G. Sykes (Eds.). *Choosing choice: School choice in international perspective* (pp. 164–195). New York: Teachers College Press.

Wan, Y. (2006). Expansion of Chinese higher education since 1998: Its causes and outcomes. *Asia Pacific Education Review, 7*(1), 19–32.

Warren, J. R., & Saliba, J. (2012). First-through eighth-grade retention rates for all 50

states: A new method and initial results. *Educational Researcher, 41*(8), 320–329.
World Bank. (2020). UIS.REPR.1.G1, UIS.REPR.1.G2, UIS.REPR.1.G3, UIS.REPR.1.G4, UIS.REPR.1.G5, UIS.REPR.1.G6. Database. World Bank Databank. 2020. https://databank.worldbank.org/home.
Yang, S.-G., & Zhu, H. (2007). Central sinking, financial weakening and studies on realizing the central rising by financial support. *Economic Research Journal, 5*, 55–67.
Yu, L., & Suen, H. K. (2005). Historical and contemporary exam-driven education fever in China. *KEDI Journal of Educational Policy, 2*(1), 17–33.
Zeng, J., Pang, X., Zhang, L., Medina, A., & Rozelle, S. (2014). Gender inequality in education in China: A meta-regression analysis. *Contemporary Economic Policy, 32*(2), 474–491.