

RESEP WORKING PAPER

Department of
Economics,
Stellenbosch
University

Working Paper No.

01/26

**MARCH
2026**

This paper was
produced as part of the
MILAPS project, funded
by Optima and utilises
Data Driven Districts
data

An unequal start:

The impact of school-entry age on grade repetition and academic performance in Grades 1 - 4

Author

Bianca Böhmer

An unequal start: The impact of school-entry age on grade repetition and academic performance in Grades 1–4

Bianca Böhmer

March 2026

Abstract

This paper estimates the effect of school-entry age on grade repetition and academic achievement in a middle-income education system characterised by high repetition, institutional flexibility in Grade 1 admissions, and relatively strong compliance with official school-entry age requirements. South African school administrative records (SA-SAMS) form the basis of a learner-level panel dataset that is used to follow a cohort of first-time Grade 1 entrants from 2018 through to Grade 4.

Although legislation permits school-entry over an 18-month window, schools effectively implement one of two de facto regimes. Either they admit learners in an earlier “Mid-year” or later “Calendar Year” age range. Because parents may also select to send their children early or delay their entry based on unobserved readiness, the realised entry age is endogenous. To address this selection, actual school-entry age is instrumented with expected school-entry age, which is calculated based on learners’ birthdays and the school-entry age regime. School fixed effects account for the lack of standardisation in outcome measures across schools.

Relatively older entrants perform substantially better in the early grades. Instrumental-variables estimates indicate that being one year older at entry reduces the probability of repeating Grade 1 by roughly 16 percentage points, relative to a mean of 13% at Mid-year schools, and by 10 percentage points for a mean repetition rate of 10% at Calendar Year schools. Older learners also achieve higher marks in home language, first additional language, and mathematics. Relative age effects are larger for boys and for learners attending poorer schools. Although these gaps attenuate, meaningful differences persist to Grade 4. Non-parametric rank-mobility analyses show that convergence is higher among lower-ranked learners in Grade 1 and is correlated with higher rates of repetition.

1. Introduction

An extensive international literature¹ shows that a child's birthday is a critical factor influencing their schooling outcomes, as a result of their age relative to their peers when they enter school. This phenomenon is known as the "Relative Age effect"², "Entrance Age effect"³ or "Birthday effect"⁴, and occurs because admissions policies generally dictate a cut-off date for school entry, determining who is able to be admitted in that year. As there is a single new intake per year, children in the same grade could be almost a year apart in age. At younger ages, such an age difference is developmentally meaningful. Relatively older learners tend to outperform their younger classmates across a range of academic and non-academic outcomes, including grade repetition, test scores and educational attainment (Bedard & Dhuey, 2006; Fredriksson & Öckert, 2005; Givord, 2020; Kawaguchi, 2011). Much of the difference in performance can be explained by the fact that relatively older learners at school entry are also relatively older at the time of assessment (Black et al., 2011; Crawford et al., 2014; Peña, 2017), but if these early experiences affect a child's motivation or shape their perceptions of their academic ability, or determine their selection into an academic track, initial disadvantages can continue to affect performance in later years (Kawaguchi, 2011; Puhani & Weber, 2007; Robertson, 2011).

Earlier work on relative age effects focused predominantly on high-income countries, addressing key measurement issues, including distinctions between absolute and relative age and between age-at-entry and age-at-test, and examining how these effects evolve into adulthood. While the existence of relative age gaps is well established, findings on when and for which outcomes these age-related advantages fade out are more system-dependent, leaving this as an open question in new contexts. More recent studies have also explored school-entry age dynamics in education systems where grade repetition is common (Diris, 2017; Givord, 2020; Jerrim et al., 2022; Pedraja-Chaparro et al., 2015), as well as documenting heterogeneous effects by socio-economic status and gender (Solli, 2017; Valdés, 2024). Evidence on relative age effects has also started to emerge from middle-income countries over the last decade, showing that such relative age effects are found in diverse settings (Peña, 2017; Platas et al., 2022; Yuret, 2018).⁵

1 This effect was documented as early as 1928 as cited in a review of 21 studies by Proctor et al. (1986). Since then, relative age effects have been identified across numerous domains, including schooling outcomes (Bedard & Dhuey, 2006; Cascio & Schanzenbach, 2016; Crawford et al., 2014; Puhani & Weber, 2007), labour market outcomes (Black et al., 2011; Fredriksson & Öckert, 2005; Kawaguchi, 2011; Nam, 2014; Zweimüller, 2013), but also areas such as sporting achievement (Sandercock et al., 2014; Wilson, 1999) or leadership (Dhuey & Lipscomb, 2008). Peña (2017) provides a more recent list of studies examining school test scores, intermediate outcomes and adult outcomes published between 2006 and 2014, summarising the methodologies employed and the core findings. Because birthdate can plausibly be assumed to be exogenous to child characteristics, expected school-entry age (based on birth date and statutory school-entry cut-offs) has also been widely used as an instrumental variable to address bias arising from confounding child characteristics when estimating the effect of actual school-entry age on a range of later outcomes (see an early paper by Angrist & Krueger (1992)).

2 Peña (2017)

3 Elder and Lubotsky (2009)

4 Usage by Peck & Trimmer (1995) was "Late birthday effect". However, this has now generally been shortened to just "Birthday effect" as in Huang & Invernizzi (2013).

5 A separate paper by McEwan & Shapiro (2008) looked at Chilean students' TIMSS scores, whilst Chile is now classified as a high-income country; at the time of publishing, it was still a middle-income country.

The South African context provides numerous reasons to examine these issues. As a middle-income country with highly unequal educational outcomes along socio-economic, racial, and gender lines (Spaull & Jansen, 2019), any additional dimensions that add to inequity in outcomes are of interest. The country also exhibits high rates of early grade repetition, with more than a quarter of learners repeating at least one grade before they enter Grade 4 (Department of Basic Education, 2023), so factors influencing repetition rates are important both from a financial and pedagogical perspective. South Africa's admissions age legislation was introduced about two decades ago and has been stably implemented over this period. However, recent legislation⁶ extended primary schooling by one year, to include a compulsory reception year (Grade R), making school-entry age questions relevant to current policy discussions. Yet surprisingly little is known about the application of the admissions policy or the consequences of school-entry age for academic progress.

This paper examines how Grade 1 school-entry age affects academic outcomes and repetition up to Grade 4 and whether school-entry age-related disparities persist or fade out over the 4 years. The paper also reviews national school-entry age patterns, assesses compliance with the official policy and describes the practices that ultimately determine when children start Grade 1. It further explores heterogeneity in relative age effects by socio-economic status and gender, thereby extending the relative age literature to another middle-income country in a context of low and unequal early grade learning outcomes.

The analysis draws on the Data Driven Districts (DDD) dataset, composed of state-collected school administrative data, to document entry-age distributions and the prevalence of under- and overage learners. It then tracks the 2018 Grade 1 cohort through to Grade 4 to determine whether there is a link between school-entry age and performance in the cohort. To correct for endogeneity in the actual school-entry age, an instrumental-variables strategy is employed. Additional non-parametric analyses, using staying probabilities and measures of upward and downward rank mobility by birth quartile, investigate fade-out and mobility patterns.

The South African evidence aligns with the relative age gaps observed internationally, finding significant relative age effects that are largest in Grade 1 (about 0.67 – 0.74 of a SD). They fade out somewhat by Grade 4 (0.34 – 0.37 of a SD), although they remain large and significant. In addition, the analysis finds that the relative age effects are larger for boys and learners from a more disadvantaged background (school Quintiles 1-3). Lastly, the results show how these relative age effects evolve between Grades 1 and 4 across the distribution, indicating that most of the "catch-up" takes place for initially under-performing learners, who are more likely to have repeated a grade.

An additional contribution of the paper is a description of how the unusual 18-month admission window in South Africa functions in practice and how this interacts with relative age effects. Schools effectively implement one of two distinct versions of the policy, either admitting children in the earliest or latest possible 12-month range. These are referred to

6 Basic Education Laws Amendment Act 32 of 2024 (BELA Bill).

as "Mid-year" and "Calendar Year" schools. The paper documents which schools are practising each policy and how they shape the age distribution of Grade 1 entrants.

There is insufficient evidence to suggest changes to existing school-entry age legislation. The legislation is already flexible, and most schools and parents comply. However, instances of early entry could be further reduced. As Grade R becomes compulsory, parents should receive clearer guidance on the benefits and potential costs of delaying entry where appropriate.

The paper proceeds as follows. Section 2 reviews the international literature on relative age effects and policy responses. Section 3 outlines South Africa's admission regulations and situates them in an international context. Section 4 describes the dataset and methodology. Section 5 presents the results, and Sections 6 and 7 discuss the findings, policy recommendations and conclusion.

2. Literature review

The Entrance Age, Birthday, or Relative Age effect describes the differences in experiences and outcomes based on a person's birthdate. This phenomenon occurs because schools follow an annual cycle, taking in one cohort of learners a year. Most countries have well-defined, often fairly prescriptive national regulations that govern the school-entry age. In combination, this means that the month of birth determines a child's relative age within their schooling cohort.

If all children of the same age start school within a year, this means that learners can be up to almost a year apart, even if they started school in the correct year. At such a young age, a 12-month period is a developmentally significant amount of time. For a child who has just turned 6 years old, a year difference in age equates to about 17% of their total lifespan. As a result, the eldest children in their cohort have a relative advantage: they are ahead in their physical, cognitive and socio-emotional development.

There is a long history of tracking differences in outcomes due to the relative age effect (Peña, 2017). The evidence across countries and using a range of methodologies is unequivocal that the eldest learners within a cohort tend to outperform their younger peers in both academic and non-academic spheres. Starting at the primary school level, older learners in a cohort tend to have lower levels of grade retention. Corman (2003) finds that in the US, children in states who just made the cut-off for school entry are 9.1 to 9.6 percentage points more likely to repeat the grade than those who missed the cut-off, whilst in France and Spain, Pedraja-Chaparro et al. (2015) show that by the age of 15, children born in the last month of the year, the youngest in their grades, are between 70% and 80% more likely to repeat a grade than their peers born in the first month. Similarly, Givord (2020) shows that across PISA countries with an average repetition rate of 7%, there is a 3 percentage points difference in repetition for learners who missed or just made the cut-off.

Oldest learners in their cohort are also found to outperform their younger peers at school (Fredriksson & Öckert, 2005) and on national and international standardised assessments

on Reading, Mathematics and Science (Bedard & Dhuey, 2006; Elder & Lubotsky, 2009; Herbst & Strawirski, 2016).

In the highly selective schooling system in Germany, children who start school in their 7th instead of their 6th year of age were 12% more likely to enter a higher secondary school (Gymnasium), on the most academic track, in their fifth year of school (Puhani & Weber, 2007). In addition, Elder & Lubotsky (2009) also show that learners in the relatively youngest birth quartile are more likely to be identified as having a learning disability⁷. In such cases, meeting academic thresholds for entry into a particular academic track or a diagnosis at a young age can strongly influence later educational trajectories. If initial entry-age advantages or disadvantages have not yet had time to fade out, this can increase the vulnerability to persistent relative age effects.

Learners who are older at the start of their schooling career relative to their peers also achieve higher levels of schooling, particularly in systems where the requirement to drop out of school is not a minimum age, but rather requires that a specific number of school years be completed (Fredriksson & Öckert, 2005; Kawaguchi, 2011), although the opposite is true in education systems that allow dropout after a minimum age has been reached (Angrist & Krueger, 1991). Bedard & Dhuey (2006) found that in Canada and the USA, learners who were born in a month that would result in them being the youngest were underrepresented by 9.8 and 11.6 percent in pre-University and four-year college enrolments.

This gap can persist into the labour market, with children who should have been the oldest at the start of their school year based on their month of birth earning higher wages in the labour market relative to those who should have been youngest in their cohort based on month of birth. In Norway, Solli (2017) finds that children born early in the year have higher earnings at age 30, whilst in Austria, Zweimüller (2013) finds an initial wage penalty of 1.1–2.0% of being the youngest; this penalty declines to 0.8–1.1% after five years of work experience. Kawaguchi (2018) tracks learners through school into the labour market in Japan. He found that older entrants achieve more years of education and also earn higher wages, but the effect on wages was only evident for men. Conversely, in Sweden, Fredriksson & Öckert (2005) find ambiguous effects on earnings, but higher initial earnings for older entrants. They attribute the eventual ambiguity to a trade-off between a later schooling start, potentially higher levels of educational attainment, and the opportunity cost of a lost year of experience. However, in the US states of Texas and California, Dobkin & Ferreira (2010) find no effect of school-entry age on labour market outcomes. Whilst in Korea, Nam (2014) finds that relative age effects persisted until middle school, but that they had faded out by the end of high school. They attribute this to the concentration and motivation of younger peers to catch up to older peers in their high school years.

Relative age effects are also evident for non-cognitive skills. Relative socio-emotional maturity, physical maturity and a more developed self-concept may also place older learners at a social advantage within their peer group. Dhuey & Lipscomb (2008) find that older students in the cohort are more likely to be selected for leadership roles such as club president or team captain, whilst other studies highlight that younger learners face a

7 Such as Attention Deficit and Hyperactivity Disorder

higher likelihood of bullying and victimisation at school (Ballatore et al., 2020; Mühlenweg, 2010). Relatedly, Black et al. (2011) show that in Norway, boys who were older at school start are less likely to suffer from poor mental health at the age of eighteen.⁸

The function of schooling, however, is about overall skills accumulation rather than achieving a particular rank with a specific cohort, leading to the important distinction between absolute age effects and relative age effects. There are two relevant absolute ages: age-at-entry and age-at-test (Altwicker-Hámori & Köllő, 2012). Age-at-entry and age-at-test are linearly related if time at school is considered. Specifically, age-at-entry plus time spent at school is equal to the age-at-test, making them perfectly collinear. Conversely, the relative age is related to the learners' position within their cohort. As a result, absolute age measures and relative age also tend to be collinear, and the effects of each can be non-trivial to disentangle.

Studies that focus on the age-at-entry as the primary measure tend to be concerned about how the absolute school-entry age affects the rate of learning or the total amount of schooling attained. The core question that is being asked is whether school entry that is delayed by a year affects learners' trajectories over the course of their schooling. On the other hand, by focusing on the age-at-test, this allows for a comparison of learners at the same absolute age. This allows for a comparison of achievement of learners at the same absolute age, as opposed to conferring an age advantage to older learners at the time of the test. However, these measures then confer an advantage to younger learners, as this group has generally been at school for a longer period of time at the time of the test.

Elder & Lubotsky (2009) use variation in US states in school entry cutoff dates to separate out the effects of absolute and relative age effects and make the case that the relative age effect is a result of the accumulation of skill prior to the start of school. But school-entry age does not appear to either positively or negatively affect a child's ability to learn at school. Their hypothesis is therefore that once a child starts formal schooling, they learn at the same rate regardless of their position within the cohort. This supports earlier findings by Morrison et al. (1997), who found that both younger and older Grade 1 learners made the same amount of progress in Grade 1, though older learners outperformed their younger peers. However, Datar (2006) found the opposite, delaying Kindergarten entry resulted in a steeper trajectory over the first two grades, and this effect was particularly strong for at-risk learners. Taking this view, the existence of a gap in the cognitive skills levels between the older and youngest learners within a cohort is then of less importance than observing how much age-dependent learning is taking place, whether there is a divergence or convergence between older and younger learners over time, or how much the rank order shifts.

A separate relative age effect is also at play: peer effects. Here, younger learners are at an advantage. They stand to gain from their interactions with older learners by directly learning skills or knowledge, or by being motivated by example. There is also reduced negative spillover from classmates who are younger and more disruptive (Cascio & Schanzenbach, 2016; Elder & Lubotsky, 2009; Peña, 2017). Specifically, Elder & Lubotsky

⁸ This effect may also exist for girls. However, the authors relied on data from the start of compulsory military service, which was collected nationally for all eligible men only.

(2009) found that children with older classmates showed improvements on standardised assessments, which are absolute measures of learning, but reduced performance on relative measures, namely, their probability of grade repetition increased, and they were more likely to be diagnosed with a learning disability. Relative age effects can also be felt through more direct means. If, for example, the instructional level is pitched at the class average, relatively younger learners could be at a disadvantage if they fall behind, whilst older learners may not be sufficiently challenged (Datar, 2006).

Lastly, assessment timing and data collection strategies directly impact what is measured. Studies that look at the performance of learners based on their age-at-test generally find that learners with a lower age-at-entry outperform learners with a higher age-at-entry. Therefore, if students were tested at the same age at different times of the year, younger students would outperform their older classmates (Black et al., 2011; Cascio & Schanzenbach, 2016; Peña, 2017).

Given that the difference of a year is a more significant difference between children at the start of their schooling than by the end of primary school, some attenuation of the relative age effect is expected over time. Several studies have found that the relative age effect is smaller in later grades and, in some cases, even fades out completely (Elder & Lubotsky, 2009; Herbst & Strawiński, 2016; Pehkonen et al., 2015; Robertson, 2011). However, in some countries, differences in some outcomes persist into adulthood in the form of differential levels of educational attainment or wages (Bedard & Dhuey, 2006; Black et al., 2011; Görlitz et al., 2022; Kawaguchi, 2011; Solli, 2017; Zweimüller, 2013).⁹

Parents are aware of these impacts and effects and have privileged knowledge about their child. Based on this knowledge, parents may choose to send their child to school early or to delay school entry. Early entry is seen as a way to give children a headstart on the investments into their education and potentially an earlier entry into the labour market; whilst delayed entry, sometimes referred to as "red-shirting"¹⁰, seeks to give children an advantage as a result of being older and more mature than their peers. Different countries and regions also follow different norms. With countries such as Ireland, United Arab Emirates and Brazil, early entry is common, whilst in places like Ukraine, Germany, the Philippines and Brazil, delayed entry is high (Givord, 2020). Overall, delayed entry is the more likely of the two.

9 There are a number of theories that attempt to explain why such a gap in performance between learners who started school a year older may persist or fade out. Three such models are explained by Robertson (2011): the student investment model, the teacher investment model and the runt of the litter model. Under the student investment model, students (and potentially parents) choose how much time and effort to invest in academics based on perceived ability. For older students, this ability would appear to be higher, resulting in higher levels of investment in schooling for older students. As a result, it is expected that the gap between older and younger students would widen over time. In contrast, under the teacher investment model, it is expected that teachers would generally cater to the average student (who is most likely to have been born in the middle of the year). As a result, the gap between the middle and oldest students would shrink and the gap between the youngest and middle students would widen. Lastly, the "runt of the litter" theory argues that relative age, rather than developmental maturity, influences academic performance. It suggests that the youngest students, due to their relative physical immaturity, may instead choose to focus on academics and consequently perform as well as or better than their older peers at school.

10 The term "red-shirting" was originally used in US college sports. Players are only eligible to play in the league for four years, and therefore holding a player back from competitive play for a season allows them to mature and train, to improve their skills and game play, whilst not using up their eligible years.

An important consideration in the debate about whether to delay entry into Grade 1 is the availability and quality of early learning alternatives available to children who delay entry into Grade 1. Morrison et al. (1997), using baseline and endline comparisons, find that learners in the US generally make greater academic progress in Grade 1 than in kindergarten (equivalent to Grade R). Similarly, Elder & Lubotsky (2009) argue that delaying school entry merely postpones the onset of more intensive human capital accumulation and therefore caution against redshirting, particularly for children from poorer households. Regarding longer-run outcomes, Black et al. (2011) find a small negative effect of later school entry on IQ test scores measured at age 18, at a time when learners are about to enter the labour market. In contrast, evidence from Hungary suggests that a one-year delay may be beneficial for children from less advantaged families (Altwicker-Hámori & Köllő, 2012). However, the authors hypothesise that this requires relatively high-quality preschool institutions, which may be better equipped than formal Grade 1 classrooms to support disadvantaged children. The provision of high-quality preschool options is likely to be more developmentally appropriate for underage learners and would provide a suitable alternative to an early start to formal schooling.

The most influential policy lever affecting school-entry age is the choice of age thresholds and permissible age ranges for entry into formal schooling. Here, the key age-related mechanism that policymakers must consider is the effect of *absolute age* on learning and development at school entry. Policy options for mitigating *relative age* effects are more limited. Internationally, two broad approaches have emerged.

The first is the use of school-readiness assessments. In the United States, such assessments were historically used to exclude age-eligible children from entering kindergarten or to place them in special kindergarten programmes (Shepard et al., 1996). This practice generally disproportionately affects poorer children, particularly when assessments are not age-appropriate or culturally sensitive. It may also be harmful where alternatives to formal schooling are of low quality or where delayed entry reduces total years of schooling completed (Barua & Lang, 2016). While the effects of redshirting remain context-dependent, some studies suggest that delayed entry may yield benefits for less school-ready children if quality pre-school or Grade R options are available (Altwicker-Hámori & Köllő, 2012; Datar, 2006).

The second approach is to "age-adjust" assessment scores in the early grades to account for age-at-test effects, particularly on standardised assessments and those used for promotion decisions. Such adjustments aim to provide a more accurate measure of learners' underlying ability and potential once differences in maturity are taken into account (Crawford et al., 2014). However, any age-adjustment approach would need to be applied carefully to ensure fairness.

3. Age of Admission Policy

The age at which children are required to begin formal schooling is generally established through national or regional legislation. Such legislation both reflects and shapes societal expectations about school readiness, child development and the organisation of education systems. The initial formulation or later adaptations to such regulations would

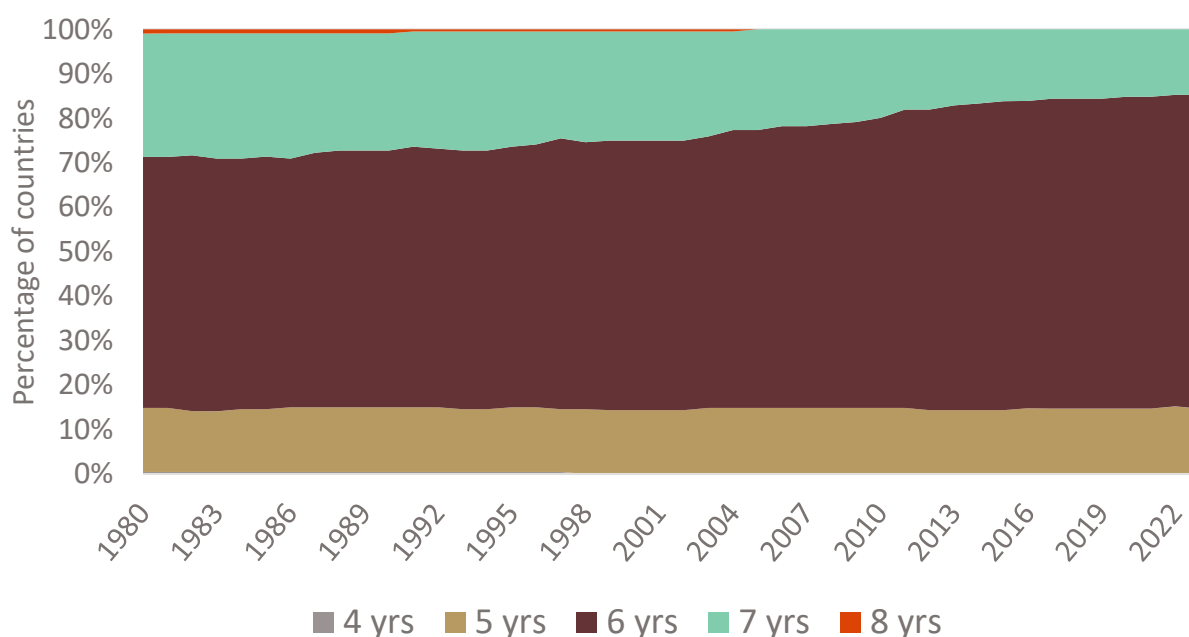
need to consider the specific country context, particularly the state of early childhood care and development. Although such regulatory changes may also be driven by political or financial considerations.

School entry regulations most commonly specify a minimum entry age, requiring that a child be admitted once they have reached a legally defined threshold, or mandate a fixed numerical age, such as six years old, at which enrolment becomes compulsory. Yet these requirements vary considerably across countries, and in some cases even between states or provinces within the same nation. Such variation shapes patterns of enrolment and determines when children first engage in structured schooling.

3.1 International comparison of primary school-entry ages

UNESCO Institute of Statistics collects and publishes comparable official primary school-entry ages across countries. Figure 1 shows how school-entry ages vary across countries and how the distribution of countries applying each school-entry age has shifted between 1980 and 2022. The most common school-entry age is six years old.¹¹ Historically, when countries have revised their admissions-age regulations, the most common change has been to lower the primary school-entry age, generally moving from seven to six years.

Figure 1: Percentage of countries with official primary school-entry age in years, 1980-2023



Source: UNESCO Institute of Statistics. Used indicator: "Official entrance age to primary education (years)" [299905]. Dataset last updated in November 2025.

¹¹ South Africa is listed as having a primary school-entry age of 7 years; this is inaccurate if viewed at the school start, so this raises a question regarding the interpretation or definition of the school-entry age in this dataset.

3.2 Age of Admission Regulation in South Africa

In South Africa, there is a slightly unusual 18-month window during which a child may enter Grade 1. The South African Schools Act (No 84 of 1996) contains two sections which regulate school-entry ages. One section specifies the minimum age of admission, whilst the other outlines compulsory school-going ages.

During the period of this analysis, from 2017 to 2023,¹² the minimum age of a learner entering Grade 1 at a public school was “age five turning six by 30 June in the year of admission.” (Section 3 in the South African Schools Act (No 84 of 1996)). The compulsory school age then sets the upper threshold of permissible school-entry ages. And, parents were required to send their children to “a school from the first school day of the year in which such learner reaches the age of seven years”.

In combination, learners were allowed to be between five-and-a-half and seven years of age on the 1st of January¹³ in the year that they are to start school. In practice, this means that children born in the first half of the year may start school when they are 5 or 6 years old, whilst children born in the second half of the year have to start school at the age of 6. A full history of the changes in school-entry age legislation is provided in the Appendix.

An 18-month school-entry age range does not fit neatly within the annual schooling cycle. Such a broad span in ages would also result in marked differences in the social and cognitive developmental stages of children within the same cohort. In practice, what is observed is that schools generally tend to admit children within a 12-month age range.

3.2.1 Two enacted policies: Mid-year and Calendar Year

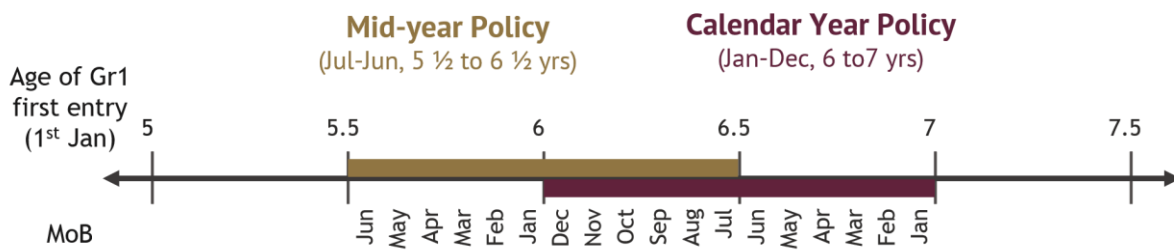
This leads to two de facto policies, which we term the “Mid-year” and “Calendar Year” policies. Under the Mid-year policy, schools admit children during the earliest possible 12-month window: learners would be between five-and-a-half to six-and-a-half years of age on the 1st of January in the year that they start school. Conversely, schools following the Calendar Year policy admit learners in the latest possible 12-month window. This includes all children between the ages of six and seven on the 1st of January in the year that they start school.

These age ranges, along with the corresponding month of birth, are shown in Figure 2. Under the Calendar Year policy, all children born within a calendar year from January to December would be part of the same cohort, whilst under the Mid-year policy, children born between July and June of the following year would be part of the same starting cohort.

12 Under the Basic Education Laws Amendment Act 32 of 2024 (BELA Bill), Grade R was included as a compulsory school year. As such, the new compulsory school age was set to the Grade R entry-age (in the year the child turns six) as opposed to the Grade 1 entry-age. However, the basic entry age range for Grade 1 would remain the same for most children.

13 The oldest possible school-entry age is for a child born on the 1st of January who turns 7 on the 1st of January in the year that they start school. The youngest possible child is one who enters Grade 1 at 5 ½ years of age, who will turn 6 on the 30th June later in the year.

Figure 2: School-entry age under the Mid-year and Calendar Year Policies



Source: Own diagram

The South African admissions policy does not stipulate that schools should restrict themselves to such a Mid-year or Calendar Year policy. However, over time, it appears as if either the Mid-year or the Calendar Year option becomes the default implemented policy at the school level.

This pattern may have originated from childcare pressures. In areas following the Mid-year policy, parents might enrol their children at school as early as possible to avoid the cost of private childcare. After the National School Nutrition Programme was introduced in 2010, the added benefit of earlier access to school meals likely strengthened this incentive.

An exception is the Western Cape¹⁴ province. In the Western Cape, the Calendar Year policy is noted as specifically preferred by the province, and they state that preference in placement is given to learners who are of the compulsory school age (Western Cape Department of Education, 2025). However, in other provinces, these practices appear to be school- or parent-determined.

4. Methodology

This paper examines whether school-entry age in Grade 1 affects South African learners' academic outcomes in up to Grade 4. Specifically, it asks:

1. How are Grade 1 admission age requirements followed?
2. How does the Grade 1 school-entry age affect Grade 1 repetition and academic outcomes?
3. How does the Grade 1 school-entry age affect FP repetition?
4. How does the Grade 1 school-entry age affect Grade 4 repetition and academic outcomes?
5. Is there a convergence or divergence in outcomes by school-entry age by Grade 4

4.1 Data

The core dataset used for this paper is derived from the Data Driven Districts (DDD) project, which consolidates school administrative data collected through the state-managed

¹⁴ The Western Cape is one of three provinces that are not included in the dataset used for the main analysis.

South African School Administrative Management System (SA-SAMS). Both public and independent schools are required to submit their data via this system.¹⁵

The dataset is an anonymised learner-level panel running from 2017 to 2023 that allows for the tracking of learners across years. It covers six of the nine provinces in South Africa: Eastern Cape (EC), Gauteng (GP), KwaZulu-Natal (KZN), Limpopo (LP), Mpumalanga (MP) and the North West (NW). These six provinces represent the large majority (over 80%) of total primary school enrolment.

A number of initial steps were taken to prepare the data. The panel dataset was checked to ensure that each learner (identified by a unique identifier) appeared only once per year, and that there were no duplicate entries across grades or schools. Learners were also tracked from grade to grade to identify whether their movement through the grades was plausible.¹⁶ In cases where it looked like two learners were erroneously assigned the same identifier (duplications), they were assigned two separate new IDs, whilst in cases where it was impossible to identify what errors had occurred, the observations were dropped. In combination, there was a negligible change (generally less than <0.5%) in the number of learners in the raw and cleaned DDD datasets in each province. A full set of comparison tables is provided in the Appendix (Table 20-Table 22).

The proportion of learners in the cleaned DDD data relative to those reported in the official School Realities Report is shown in Table 1. Within the six provinces, the DDD dataset covers about 90% of schools and learners in both public and independent schools. There is lower coverage in Gauteng and KwaZulu-Natal, where just over 80% of learners were in the data. Coverage was also lowest in 2017, as this was just after the transition to the new EMIS system.

Table 1: The proportion of learners in the cleaned DDD data relative to enrolment reports in School Realities for Grades 1-4 in 2017 to 2023

| Year | EC | GP | KZN | LP | MP | NW | All 6 |
|---------------------------|------------|------------|------------|------------|------------|------------|--------------|
| 2017 | 99% | 68% | 56% | 93% | 95% | 96% | 79% |
| 2018 | 98% | 86% | 73% | 99% | 101% | 93% | 89% |
| 2019 | 99% | 90% | 90% | 99% | 98% | 97% | 94% |
| 2020 | 98% | 81% | 82% | 97% | 96% | 93% | 89% |
| 2021 | 100% | 88% | 84% | 99% | 98% | 94% | 92% |
| 2022 | 100% | 87% | 88% | 99% | 98% | 97% | 93% |
| 2023 | 99% | 83% | 89% | 99% | 97% | 91% | 91% |
| Total | 99% | 83% | 80% | 98% | 97% | 95% | 90% |
| <i>Total (excl. 2017)</i> | 99% | 86% | 84% | 98% | 98% | 94% | 92% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GP, KZN, LP, MP and NW and School Realities Reports 2017 – 2023.

¹⁵ All provinces except the Western Cape use SA-SAMS. The Western Cape had set up its learner management system, CEMIS, before the introduction of SA-SAMS. Similar information is collected and stored on the CEMIS system as on SA-SAMS, with the two systems running in parallel.

¹⁶ For example, not skipping more than one grade or moving backwards one or more grades from one year to the next.

For the analyses in this paper, we track the cohort of first-time Grade 1 entrants in 2018 through to Grade 4. These two grades, Grade 1 and Grade 4, are specifically chosen as they are the first within their respective Phases, the Foundation Phase and the Intermediate Phase. Learners in South Africa are only allowed to repeat a grade once a Phase; as such, the first time that a Grade 1 repeater would be eligible to repeat would be in Grade 4.¹⁷ Therefore, Grade 4 is the next grade after Grade 1 that provides an unrestricted and comparable measure of repetition.

4.1.1 Balanced panel of the 2018 Grade 1 cohort

The 2018 cohort was chosen for the analysis as this was the first year in which learners who entered Grade 1 for the first time could be identified, because repeaters from 2017 could be tracked and excluded. Starting in 2018 also means that the cohort completed two full years of school before the onset of the COVID-19 pandemic, which interrupted instruction and also resulted in changes to both policy and practice relating to learner repetition.

As fewer schools submitted data in 2017, a balanced panel of schools was created. Only learners who attended Grade 1 at a school with Grade 1 data in 2017 and 2018 and Grade 4 data in 2021 and 2022 were included. This was done to ensure that Grade 1 first-time entrants in 2018 could be correctly identified from the 2017 data, provided the child remained within the same school or moved to another school within the sample.¹⁸ Note that learners who moved between schools were included, provided they moved to another school that was also included in the school panel. Table 2 shows that we drop about 150 000 learner observations (17.7%) during this step, largely due to lower school coverage in 2017.

Table 2: Number of Grade 1 learners in 2018 in various datasets

| Description | Number of 2018 Grade 1 observations |
|--|--|
| Enrolment in School Realities | 965 097 |
| Raw DDD data | 859 161 |
| Cleaned & ordered DDD data | 860 300 |
| School Panel (Gr 1 – 2017/18, Gr 4 2021/22) | 708 403 |
| Grade 1 & 4 learner balanced panel (incl. repeaters) | 633 482 |
| Grade 1 first-time entrants only balanced panel | 556 562 |

Then, a balanced panel was created by restricting the sample to the learners from the 2018 Grade 1 cohort who could be tracked from Grade 1 to Grade 4. This resulted in the loss of approximately 75 000 (10.6%) observations. Lastly, we excluded all learners who were repeating Grade 1 in 2018, which amounted to about 77 000 (~12.1%) observations.

As comparable information is only available within a grade level regardless of the age of the learners, we use a same-grade approach. The outcomes that are used for each learner are taken from the first year in which they attempted the grade level. This means that in

¹⁷ Grade 4 is the first year of the Intermediate Phase (Grades 4-6). The other two Phases are the Senior Phase (Grades 7-9) and finally the Further Education and Training Phase (Grades 10 – 12)

¹⁸ Specifically, this approach aims to account for cases where a school appears in the 2018 dataset but not in the 2017 dataset. Here, all learners in the 2018 Grade 1 cohort would incorrectly be assumed to be first-time entrants, since repeating learners who attended Grade 1 in 2017 would not be identified as having repeated.

Grade 4, the outcomes are from different calendar years depending on whether learners had repeated in the Foundation Phase, and the age-at-test is at least one year older for repeaters.

Table 3 shows the final sample of 556 562 learners and the years in which they show up in Grade 1 and in Grade 4. About 75% of learners make it to Grade 4 by 2021 without repeating a grade. The provincial split of learners in the sample is shown in Table 4.

Table 3: Number of learners in the final balanced panel of the Grade 1 2018 first-time entrants tracked to Grade 4, showing repeaters

| Grade | Year | | | | | |
|--------------------------|---------|--------|-------|---------|---------|--------|
| | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| With repeaters | | | | | | |
| 1 | 556 562 | 70 527 | 1 158 | - | - | - |
| 4 | - | - | 452 | 419 983 | 139 540 | 30 518 |
| Without repeaters | | | | | | |
| 1 | 556 562 | - | - | - | - | - |
| 4 | - | - | 452 | 419 800 | 122 734 | 13 576 |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW.

Table 4: Number of learners in the final balanced panel of the Grade 1 2018 first-time entrants by province

| | Province | | | | | | Total |
|------------------------------|----------|---------|--------|---------|--------|--------|----------------|
| | EC | GT | KZN | LP | MP | NW | |
| Learners in Gr 1 2018 Cohort | 115 273 | 109 347 | 86 930 | 118 685 | 72 110 | 54 217 | 556 562 |
| <i>% of Sample</i> | 21% | 20% | 16% | 21% | 13% | 10% | 100% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW.

4.2 Analytical approaches

4.2.1 Measuring school-entry age compliance

A simple count was conducted to assess compliance with the Grade 1 school-entry age requirements. Learners were grouped according to their age on the 1st of January in the year they first entered Grade 1. To correctly identify first-time Grade 1 entrants, we needed to be able to detect learners who were repeating Grade 1. This requires that repeaters at the same school can be tracked, which is not possible if the school did not submit data in the previous year. For this reason, a school was included in the sample for year X only if learner data for that school were available in the preceding year (X-1). This rolling approach allowed us to generate an estimate each year while retaining as much data as possible. Although it may introduce some incomparability, particularly between 2018 and later years due to lower data completeness in 2017, it remains preferable to using a true balanced

panel. Restricting the dataset to only schools present in every year would remove a substantial amount of information, improving comparability across years but increasing bias in the reported measures for all years except 2018.

4.2.2 Effect of school-entry age on academic outcomes

For research questions 2, 3 and 4, the analysis aims to determine whether, and to what extent, children's academic performance differs from that of their peers based on their relative school-entry age. The main analytical framework can be summed up in the following equation, which utilises an approach taken by Angrist & Krueger (1992) and expanded by Bedard & Dhuey (2006) and Elder & Lubotsky (2009):

$$Y_i = \alpha + \beta AE_i + \lambda \mathbf{X}_i + \varepsilon_i$$

Here Y_i is the end-of-year academic outcome for learner i . Four different outcomes are tracked: repetition, and school-based assessment (SBA) marks in Home Language (HL), First Additional Language (FAL) and Mathematics (MTH)

In the equation AE_i is the age at which learner i first enters Grade 1, and \mathbf{X}_i is a vector of covariates, including the learner's gender, whether the learner attended Grade R at a school, the school Quintile¹⁹ (a wealth measure) of the school attended, whilst ε_i is the error term. The AE_i is likely to be endogenous due to omitted variable bias. Parents can choose to send their child to school a year early or a year late. If this choice is influenced by unobserved child characteristics, the actual school-entry age will be endogenous. These unobservable characteristics, such as a child's academic ability, social maturity, health, or other family or community factors, are confounders: they influence both the decision to enrol a child early or late and also influence later schooling outcomes. To address this, the expected school-entry age, determined by date of birth relative to the school-entry age cutoff, provides a source of plausible exogenous variation. Conditional on school fixed effects, which account for socio-economic and regional variability, birth timing should be unrelated to individual ability.

¹⁹ In South Africa, school Quintiles is an official designation used to determine the funding that a school receives—they do not correspond to true quintiles. Schools in the lowest three school Quintiles are not allowed to charge school fees and receive the highest amount of per learner funding (excluding teacher costs). School Quintiles were originally assigned based on the poverty level of the area in which a school is located. However, over time, the wealth status of various areas has shifted, and schools have campaigned to be reassigned to lower quintiles to access higher levels of funding. As a result, the lower quintiles comprise more than 60% of schools and learners, and Quintile 5 has less than a 20% share.

The Wu-Hausman test of endogeneity²⁰ was first conducted to assess the endogeneity of school-entry age due to the omitted variable bias (Hausman, 1978). The test used Grade 1 repetition as the outcome variable and expected school-entry age as the instrument. It produced an F-statistic of 1 260 with a p-value of 0.0000, indicating a very high likelihood that actual school-entry age is endogenous. When the sample was analysed separately, restricted to only Mid-year or Calendar Year schools, the F-statistics were lower, but the p-values remained effectively zero (up to five decimal places). Based on these results, an instrumental variables (IV) approach was applied.

In order to correct for this bias, using the expected school-entry age as an instrument is explored. For an IV (z) to be a valid instrument, two conditions of relevance ($Cov(z, AE) \neq 0$) and exogeneity ($Cov(z, \varepsilon) = 0$) need to be met.

For the IV to be relevant, it needs to be correlated with the variable AE_i , but also uncorrelated with the error term. The first condition is clearly met, as illustrated in Figure 3. In addition, the F-statistic of the first stage with no controls was 1.4×10^5 for Mid-year schools and 18 398.61 for Calendar Year schools, indicating strong instrument relevance.

Figure 3: Mean Grade 1 first-time entry age by month-of-birth and Mid-year or Calendar Year policy



Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW.

Using the birthdate or expected entry-age as an instrument is a standard approach in the literature (Bedard & Dhuey, 2006; Black et al., 2011; Cascio & Schanzenbach, 2016; Datar, 2006; Elder & Lubotsky, 2009; Fredriksson & Öckert, 2005; Puhani & Weber, 2007), which requires that a child's birthday is not correlated with unobserved child characteristics.²¹

20 The Durbin-Wu-Hausman test of endogeneity (χ^2 -test) produces an almost identical result.

21 Other approaches have included a regression discontinuity around the age thresholds (McEwan & Shapiro, 2008; Valdés, 2024), using date of birth as a proxy for relative age, more often used in education systems

However, the approach has raised objections for a number of reasons.²² A key criticism is that annual birth patterns are not random (Bound & Jaeger, 1996). Firstly, the likelihood of pregnancy is not necessarily evenly spread across the year. Demographic factors such as class, race, region and mother's age have been shown to be correlated with seasonal birth patterns (Bound & Jaeger, 2000; Buckles & Hungerman, 2013). Additionally, annual events such as holidays have an impact on sexual activity (Symul et al., 2022), whilst explicit parental choice to time their pregnancies may also influence birthdates. A second concern is that seasonal health and environmental factors influence foetal, birth and infant outcomes, including low birthweight, gestational period, miscarriages and stillbirths, but may also have longer-run effects that influence schooling outcomes.²³ Separately, Barua & Lang (2016) show that using the expected school-entry age as an IV can be inconsistent if it does not meet the monotonicity condition. Lastly, the sample size remains a consideration when using IVs (Bound et al., 1995).

In Sub-Saharan Africa, there is some birth seasonality; however, it tends to have weaker periodicity relative to other regions or fluctuations over time (Dorélien, 2016). This muted seasonality is desirable for using the birthdate as an IV. National birth registration data in South Africa from 2008 to 2017 show some seasonality, with a large peak in September (+7.6%), a smaller peak in March (+4.6%) and a defined trough in October (-9.1%) and November (-8.3%).²⁴ The likelihood of being born in the first six months of the year over these ten years was slightly higher, at about 50.8%, despite the first half of the year only containing 49.6% of the days (see Figure 13 in the Appendix). In the DDD dataset, the incidence of birthdays by month shows slight differences across provinces, language and race groups in the 6 provinces of interest. However, such differential seasonality is likely to apply similarly to children born within the same geographic area and having a similar demographic and socio-economic background. The month of birth is less likely to be correlated to the characteristics and abilities of a specific child from that group once group-specific factors have been accounted for, although environmental determinants through a health channel may still play a role. Seasonality is addressed in two ways. Firstly, school fixed effects are included, which effectively control for regional and socio-economic patterns in birth seasonality. Secondly, there are two possible cohorts that learners in South Africa can belong to, with birthday ranges six months apart. If similar patterns are found for each of the cohorts, then the likelihood that seasonality is driving the result is reduced.

Measurement considerations

Additionally, two further factors need to be considered in the regression analysis. The first relates to the school-entry age policy for South African schools. Despite the official 18-

where delayed entry is extremely rare or not allowed and grade retention is rare (Kawaguchi, 2011), and a difference-in-difference approach using variation in age thresholds over time or space (Peña, 2017).

22 Many of the initial objections focus on the use of the IV to estimate educational attainment and longer-run adult outcomes as opposed to schooling outcomes.

23 This includes factors such as ambient temperature, sunshine and vitamin D exposure, air pollution, infectious diseases, food availability and nutrition and physical activity (Arroyo et al., 2016; Dorélien, 2016; Strand et al., 2011).

24 The September and March peaks correspond to the summer and winter school holidays. December (colloquially referred to as Ke Desemba) is a time of celebration where families gather, and when individuals who travel or migrate for work are more likely to be at home. Whilst the start of the year tends to be busy and many families experience financial strain (Janu-worry), this may result in reduced sexual activity.

month Grade 1-entry window, in practice, two distinct policy options are enacted: Mid-year or Calendar Year. Because these result in different default school-entry age ranges, schools following each policy are analysed separately.

The second factor is measurement error that is inherent in the school term marks used in the analysis. The school assessments from which the marks are derived are designed and administered at the school level. As a result, there is variation in the standard of assessment across schools. To account for these school-level differences, school fixed effects are included in the regressions.

4.2.3 Identifying fadeout from Grade 1 to Grade 4

The measure of fadeout is a relative one; it asks whether a learner in Grade 4 is performing better or worse than they were in their Grade 1 year, relative to the rest of their cohort. In order to calculate this at the school level, percentile ranks were created for each learner in Grade 1 and Grade 4 within their school. This was done, partly, because this is a more robustly comparable measure across schools and across years, given the high rates of repetition across the Foundation Phase and the lack of standardisation in school marks across schools and across years. We first assess changes in the percentiles using a two-stage least squares regression with expected school-entry age as an instrumental variable, as described in the section above.

To compare the transition of members of the youngest and oldest birth quartiles of the Grade 1 cohort over the course of the Foundation Phase, a non-parametric approach is used. Bhattacharya & Mazumder (2011) introduced conditional rank mobilities to track black-white differences in intergenerational income mobility in the US, and these were also employed by McDonough (2015) to track changes in academic achievement gaps.

These transition metrics provide a more detailed look at transitions taking place over the Foundation Phase. These are used to track changes in learner percentile rank within their school from Grade 1 to Grade 4, depending on their birth quartile.

Staying probability

Following the approach by McDonough (2015), learners are tracked from an earlier grade (t_0) into a later grade (t_1), with the learner's schooling outcomes in each grade, respectively, represented by Y_{t_0} and Y_{t_1} . The overall distribution of school outcomes in each grade is represented by the cumulative distribution functions $F_{t_0}(\cdot)$ and $F_{t_1}(\cdot)$.

Then, for a defined quantile range with lower (k_1) and upper (k_2) thresholds with $0 \leq k_1 < k_2 \leq 1$, the probability that a learner remains within the defined quantile range in the later grade, conditional on them having been in that same quantile range in the previous grade, is given by:

$$\psi_{t_0,t_1} = \frac{\Pr[k_1 \leq F_{t_1}(Y_{t_1}) \leq k_2 , k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2]}{\Pr[k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2]}$$

Furthermore, this approach allows for the calculation of a direct difference conditioning on a set of covariates X , which becomes:

$$\psi_{t_0,t_1} = \int \Pr[k_1 \leq F_{t_1}(Y_{t_1}) \leq k_2 \mid k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2, X = x] dF(x \mid k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2)$$

and allows for the comparison of staying probabilities for learners from different quarters of birth.

Upward and Downward Rank Probabilities

Using the same definitions as above, for an upward shift denoted by $\delta \in [0, 1 - s_2]$ and $0 < s_1 < s_2 < 1$ to ensure that observations with unachievable upward mobility are not included. The probability that a learner improves their quantile ranking by δ from the earlier to the later grade is given by:

$$\theta_{t_0,t_1}^{\delta} = \frac{\Pr[F_{t_1}(Y_{t_1}) - F_{t_0}(Y_{t_0}) > \delta, k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2]}{\Pr[k_1 \leq F_{t_0}(Y_{t_0}) \leq k_2]}$$

The same logic is applied to obtain the Downward Rank Probability by allowing $\delta \in [0 + s_1, 1]$ and including a negative sign to the delta in the first statement of the main equation, and reversing the inequality.

5. Results

Given the presence of two distinct default school-entry age policies, this section outlines the characteristics of the full sample and then identifies how these two groups differ.

5.1 Summary statistics

An overview of the sample characteristics is presented in Table 5. The summary statistics are broken out by the default school-entry age policy. The average school-entry age of learners at Mid-year schools was about 6.07 years of age, whilst at Calendar Year schools, the average age on 1 January in their Grade 1 year was 6.53 years of age. This aligns well with the expected midpoint of the age ranges. School marks across the three subjects, HL, MTH and FAL, were slightly lower on average in Mid-year schools than in Calendar Year schools. The repetition rate in Grade 1 is high at 13% and almost twice as high over the course of the Foundation Phase (FP) (Grades 1-3) as a whole at 25%, with a lower repetition rate of 7% in Grade 4. In all four grades, repetition rates were lower in Calendar Year schools.

However, these differences in repetition rates and school marks need to be interpreted in context. Calendar Year schools tend to be wealthier, proxied by the school Quintile, and are more common in some of the provinces, including the most urban province, Gauteng, and the North West. Such school differences will influence learners' academic outcomes.

Table 5: Table of summary statistics by applied school-entry age policy, 2018 Grade 1 cohort panel

| Variable | Applied school-entry age policy | | Total |
|---|---------------------------------|---------------|-------|
| | Mid-year | Calendar Year | |
| Age in Gr 1 (On 1 st Jan) | 6.07 | 6.53 | 6.12 |
| HL mark (%) | 65.22 | 67.19 | 65.47 |
| MTH mark (%) | 69.50 | 72.09 | 69.81 |
| FAL mark (%) | 66.15 | 69.15 | 66.52 |
| Repeated Gr 1 | 0.13 | 0.10 | 0.13 |
| Repeated at least once in FP | 0.25 | 0.18 | 0.25 |
| Repeated Gr 4 | 0.07 | 0.05 | 0.07 |
| Province | | | |
| EC | 0.20 | 0.28 | 0.21 |
| GP | 0.19 | 0.27 | 0.20 |
| KZN | 0.17 | 0.04 | 0.16 |
| LP | 0.24 | 0.04 | 0.21 |
| MP | 0.13 | 0.16 | 0.13 |
| NW | 0.08 | 0.20 | 0.10 |
| School Quintile | | | |
| Q1 | 0.29 | 0.03 | 0.26 |
| Q2 | 0.27 | 0.07 | 0.25 |
| Q3 | 0.28 | 0.23 | 0.27 |
| Q4 | 0.09 | 0.16 | 0.10 |
| Q5 | 0.05 | 0.43 | 0.09 |
| Missing Q. | 0.00 | 0.01 | 0.00 |
| Independent | 0.02 | 0.07 | 0.03 |
| Home Language subject²⁵ | | | |
| Afrikaans | 0.01 | 0.25 | 0.04 |
| English | 0.13 | 0.51 | 0.17 |
| IsiNdebele | 0.01 | 0.00 | 0.01 |
| IsiXhosa | 0.18 | 0.07 | 0.17 |
| IsiZulu | 0.22 | 0.05 | 0.20 |
| Sepedi | 0.17 | 0.00 | 0.15 |
| Sesotho | 0.03 | 0.01 | 0.03 |
| Setswana | 0.10 | 0.09 | 0.10 |
| SiSwati | 0.04 | 0.00 | 0.04 |
| Tshivenda | 0.04 | 0.00 | 0.04 |
| Xitsonga | 0.06 | 0.00 | 0.05 |
| SASL ²⁶ | 0.00 | 0.00 | 0.00 |

25 In the Foundation Phase the HL subject is usually also the language of learning and teaching (LOLT) at the school. Often, it is also the actual home language of most of the learners at the school. However, in urban areas (especially Gauteng Province), learners' home languages are less likely to match the language of their home-language subject.

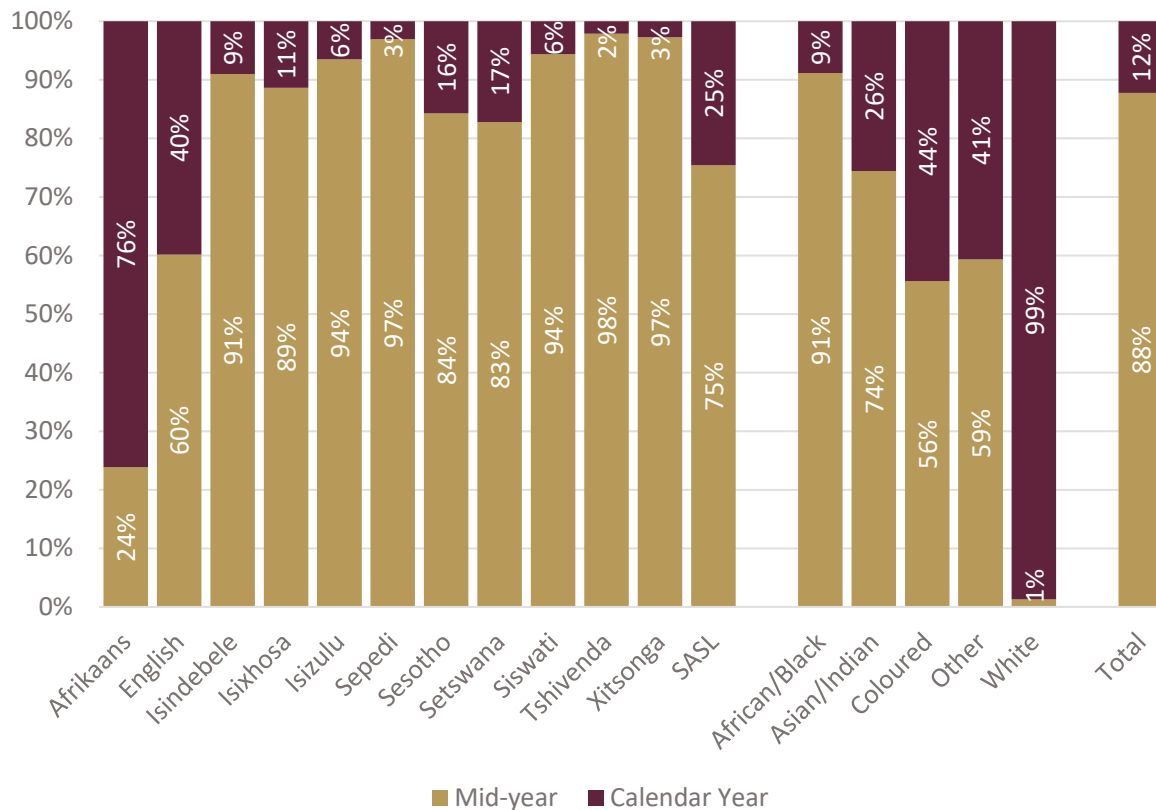
26 SASL is an abbreviation for South African Sign Language.

| Variable | Applied school-entry age policy | | |
|----------------------|---------------------------------|---------------|----------------|
| | Mid-year | Calendar Year | Total |
| Home Language | | | |
| Afrikaans | 0.01 | 0.25 | 0.04 |
| English | 0.03 | 0.13 | 0.04 |
| IsiNdebele | 0.02 | 0.01 | 0.01 |
| IsiXhosa | 0.20 | 0.18 | 0.20 |
| IsiZulu | 0.25 | 0.13 | 0.24 |
| Sepedi | 0.18 | 0.04 | 0.16 |
| Sesotho | 0.04 | 0.05 | 0.04 |
| Setswana | 0.10 | 0.15 | 0.10 |
| SiSwati | 0.05 | 0.02 | 0.04 |
| Tshivenda | 0.05 | 0.01 | 0.04 |
| Xitsonga | 0.07 | 0.01 | 0.07 |
| SASL ²⁶ | 0.00 | 0.00 | 0.00 |
| Race | | | |
| African/Black | 0.97 | 0.68 | 0.94 |
| Asian/Indian | 0.01 | 0.02 | 0.01 |
| Coloured | 0.02 | 0.10 | 0.03 |
| Other | 0.00 | 0.00 | 0.00 |
| White | 0.00 | 0.20 | 0.03 |
| Observations | 488 492 | 68 070 | 556 562 |
| % | 87.8% | 12.2% | 100.0% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy.

Figure 4 shows the split of learners by school-entry age policy type for each home language and race group for the Grade 1 2018 cohort. Schools with a high proportion of Afrikaans and English home language learners are more likely to be following a Calendar Year policy, while learners who have Sepedi, Tshivenda, and Xitsonga as their home languages almost exclusively attend Mid-year schools. There is also a clear split by race, with 91% of African Grade 1 learners going to Mid-year schools, whilst only 1% of White learners in this sample do so. Although it is important to note that a significant share of White, and Indian or Asian learners were underrepresented in GT and KZN in the data for 2017. Therefore, the 2018 Grade 1 cohort sample is not representative of these groups or provinces (see Table 24 in the Appendix). Nevertheless, this suggests that school-entry age policies may also be correlated with other unobservable factors, even for schools that are at similar wealth levels.

Figure 4: Percentage of Grade 1 learners by home language and race attending schools by each school-entry age policy type in 2018



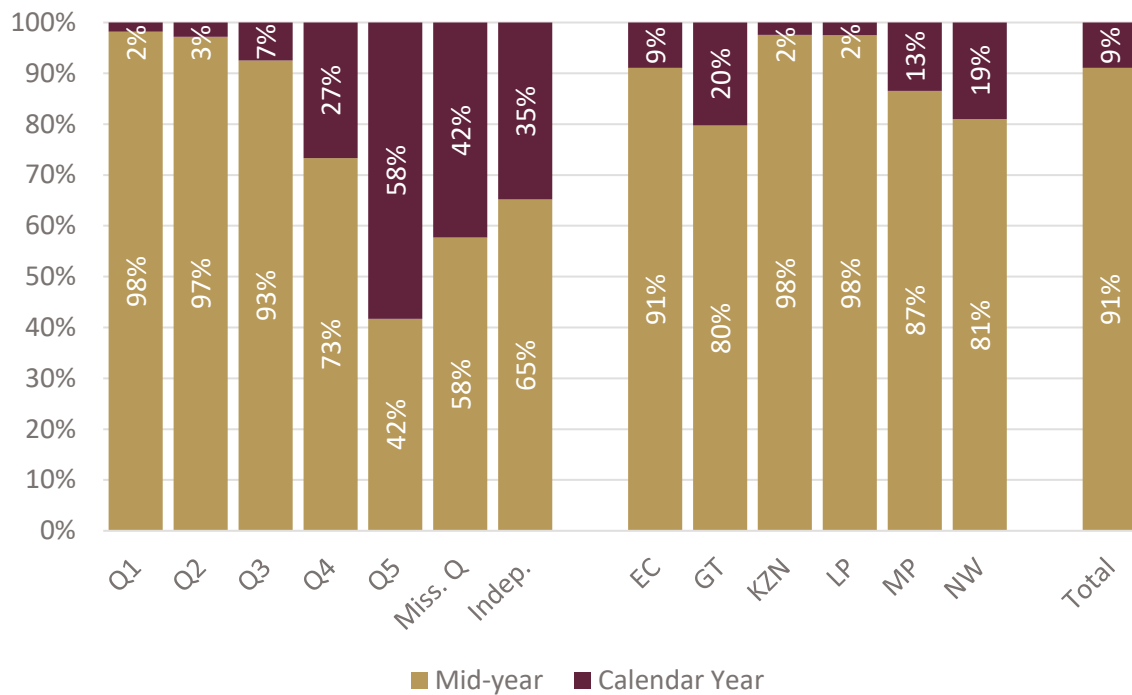
Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy.

A school-level view is presented in Figure 5, where the proportion of schools following each of the school-entry age policies is shown by school Quintile and province. Wealthier schools are more likely to follow a Calendar Year policy, and poorer schools tend to follow a Mid-year policy. Almost all Quintile 1 and 2 (~97–98%)²⁷ and most Quintile 3 schools (~93%) practice the Mid-year policy, whilst more than half of Quintile 5 schools and about a third of independent schools follow a Calendar Year policy. Different school Quintile distributions within the provinces partially determine the provincial differences; however, there also appear to be regional norms that influence which age range is used. In Limpopo and KZN, even Independent and Quintile 5 schools tend to follow the Mid-year policy, resulting in an earlier school start, whilst in the North West and Gauteng, a Calendar Year start is more common, even in lower Quintile schools.

²⁷ The default school-entry age policy was assigned based on the ages of the learners entering Grade 1 at the school. It is possible that the 2–3% of schools that are identified as Calendar Year schools in Quintiles 1 and 2 are schools that would otherwise have been Mid-years schools that simply have a large proportion of overage learners, which may be a signal of school or community dysfunction.

This pattern of earliest possible entry likely developed due to childcare constraints and, later, access to school meals in poorer communities. These pressures would have been less pronounced in wealthier Quintile 5 and independent schools. Views on the benefits of an early school start are also likely to be less prevalent among wealthier families; instead, some may even have a preference to “red shirt”. As a result, opting for a later start may have been preferred in this setting.

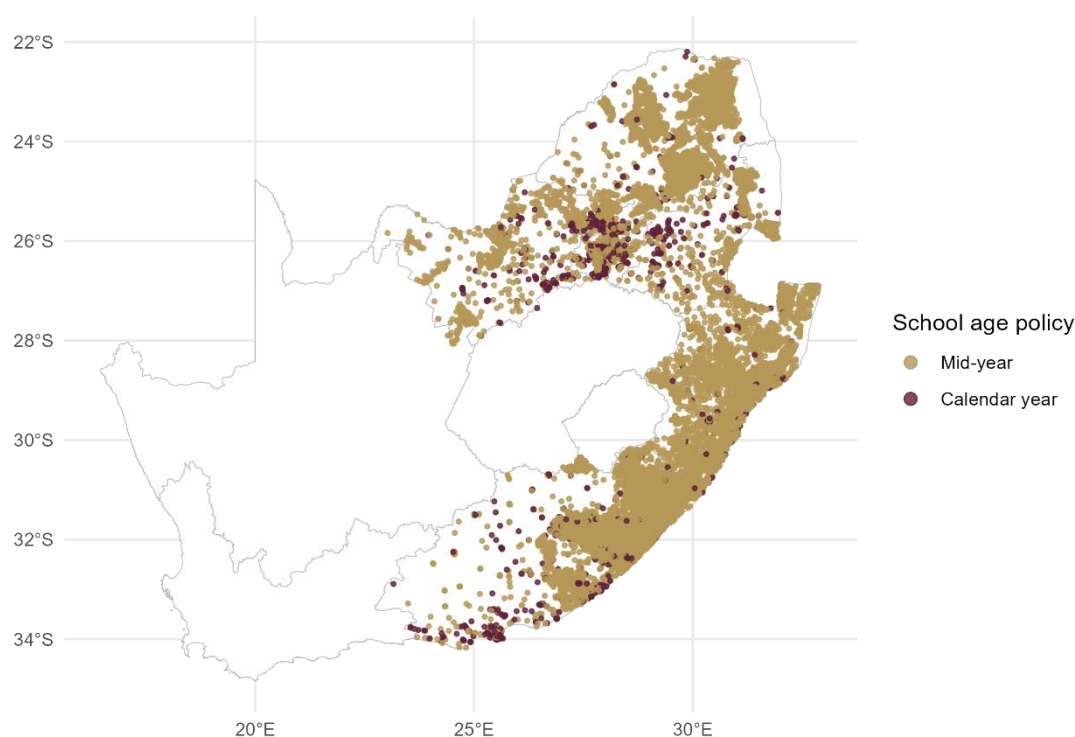
Figure 5: Percentage of schools adhering to each school-entry age policy type, by school Quintile and province in 2018



Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Schools were assigned to each policy (Mid-year or Calendar Year) based on the proportion of learners in a given year whose ages fell within the relevant 12-month age range in the years 2018 - 2023. The most commonly used policy by that school over time was then assigned to each school. Number of schools = 12 099

The geographic spread of schools is shown in Figure 6. Most schools are Mid-year, with Calendar Year schools clustered in Gauteng, in the southeast of the North West province, western Mpumalanga, and coastal cities of the Eastern Cape.

Figure 6: Map of primary schools showing the school-entry age policy followed by each school in 2018



Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Schools were assigned to each policy (Mid-year or Calendar Year) based on the proportion of learners in a given year whose ages fell within the relevant 12-month age range in the years 2018 - 2023. The most commonly used policy by that school over time was then assigned to each school. Number of schools = 12 099.

5.2 Age compliance

Compliance with the school-entry age to Grade 1 in South Africa is fairly high by international standards (Givord, 2020). However, this is when South African compliance is being compared to systems where a 12-month entry-age range is the norm.

Table 6 shows the proportion of learners who entered school between 2018 and 2023 who were under-, over- or correctly aged according to the legislated 18-month age range. Compliance appears to have increased somewhat between 2018 and 2020, although the 2018 numbers should be treated with caution as the proportion of schools that submitted data in 2017 is much lower than in subsequent years, reducing the number of schools for whom compliance can be measured in 2018.

An increase in compliance is not unexpected. The SA-SAMS system became the default system for submitting learner data in 2017²⁸, whilst the DDD project's national roll-out took place in 2016 and 2017²⁹. This electronic learner and school management system, and

²⁸ School switched from submitting learners and school data through the SNAP survey to doing so via SA-SAMS. The last year for the SNAP survey was 2016 (<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/689>). A SA-SAMS timeline is shown here: <https://sa-sams.nect.org.za/sa-sams-timeline/>

²⁹ <https://www.dell.org/ideas/data-driven-districts/>

consolidated dashboard would have allowed for greater oversight, which is likely to have contributed to some improvement in compliance over time. The very slight increase in overaged learners starting Grade 1 in 2021 and 2022 is likely a result of some parents having decided to delay school entry as a result of the pandemic. But by 2023, levels of overage entry were back to 2020 levels at around 3.0%.

Table 6: Over-, under- and correctly-aged first-time grade 1 entry, 2018 to 2023

| | Year of Grade 1 first-time entry | | | | | |
|--------------------|----------------------------------|----------------|----------------|----------------|----------------|----------------|
| | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| Underage | 2.6% | 2.0% | 1.3% | 1.1% | 1.3% | 1.4% |
| Correct age | 93.7% | 94.7% | 95.7% | 95.2% | 94.8% | 95.6% |
| Overage | 3.7% | 3.3% | 3.0% | 3.6% | 3.9% | 3.0% |
| Sample size | 651 567 | 725 739 | 727 017 | 721 846 | 713 455 | 699 554 |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Only included schools that had information in the previous year, to ensure that repeaters and first-time entrants into Grade 1 could be correctly identified, provided the child stayed within the same school.

If over- and underage entrants are broken down by month of birth, a clear pattern emerges around the minimum and maximum age thresholds. Table 7 shows compliance patterns by month of birth. Parents of children born in the first half of the year have the option of sending their child to school when they are 5 or 6 years old; however, children born in the second half of the year have to start school when they are 6 years old. As a result, compliance is very high (97.8%) for children born in the first half of the year, but substantially lower (89.4%) for children born in the second half of the year.

Table 7: Over-, under- and correctly-aged first-time grade 1 entry by month of birth, 2018

| | Month of Birth | | | | | | |
|--------------------|----------------|-------|-------|-------|-------|-------|-----------------------|
| | Jan | Feb | Mar | Apr | May | Jun | 1st Half (Jan-Jun) |
| Underage | 0.6% | 0.4% | 0.3% | 0.2% | 0.2% | 0.2% | 0.3% |
| Correct age | 97.9% | 97.8% | 97.9% | 97.8% | 97.6% | 97.8% | 97.8% |
| Overage | 1.6% | 1.8% | 1.8% | 1.9% | 2.2% | 2.0% | 1.9% |
| | Jul | Aug | Sept | Oct | Nov | Dec | 2nd Half (Jul-Dec) |
| Underage | 10.8% | 5.9% | 4.8% | 3.7% | 2.7% | 2.3% | 5.1% |
| Correct age | 85.4% | 89.7% | 90.4% | 90.8% | 90.7% | 89.5% | 89.4% |
| Overage | 3.8% | 4.4% | 4.8% | 5.5% | 6.6% | 8.2% | 5.5% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Only included schools that had information in the previous year, to ensure that repeaters and first-time entrants into Grade 1 could be correctly identified, provided the child stayed within the same school. Sample size = 651 566

Regulations allow for an 18-month school-entry age window, which is used in Table 7. However, schools and parents typically align with either a Mid-year or Calendar Year admissions policy. Table 8 shows the compliance with the narrower Mid-year and Calendar Year school-entry age policies. Only 6.7% of learners at Mid-year schools complied with

the national policy, but did not follow the norm at their school; similarly, only 4.9% of learners at Calendar Year schools started earlier than normal for their school, despite being allowed to do so under the national policy.

These patterns support the existence of a preferred choice of school-entry age that varies by region or school. Schools may be actively encouraging parents to enrol children within one of two 12-month periods rather than across the full legislated 18-month window, or such patterns may have developed as a result of parental preferences that then became institutionalised. But the evidence indicates substantial and systematic variation in how the national policy is implemented at the local and school level.

Table 8: Table to show the school-entry age compliance rates, by school-entry age policy in 2018

| | Mid-year | Calendar Year |
|-------------------------|----------|---------------|
| Less than 5yr6mo | 3.0% | 0.3% |
| 5yr6mo - 5yr11mo | 87.2% | 4.9% |
| 6yr0mo - 6yr5mo | | 87.4% |
| 6yr6mo - 6yr11mo | 6.7% | |
| 7yr0mo or older | 3.1% | 7.4% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Only included schools that had information in the previous year, to ensure that repeaters and first-time entrants into Grade 1 could be correctly identified, provided the child stayed within the same school. Sample size = 651 566 (MY = 569 862 and CY = 81 704)

There are also clear gender differences in the likelihood of being under- or overaged. Among learners entering school at the correct age, the gender distribution is close to even: in Mid-year schools, 50.1% of correctly-aged learners are boys, and in Calendar Year schools, the figure is 49.6%, as seen in Table 9. However, underaged learners are more likely to be girls, whereas over-aged learners are much more likely to be boys. For learners over the age of 7 at school entry, almost 60% are boys.

Table 9: Proportion of boys in each school-entry age group, by Mid-year and Calendar Year schools

| | Mid-year | Calendar Year |
|-------------------------|----------|---------------|
| Less than 5yr6mo | 45.8% | 41.7% |
| 5yr6mo - 5yr11mo | 50.1% | 46.3% |
| 6yr0mo - 6yr5mo | | 49.6% |
| 6yr6mo - 6yr11mo | 55.8% | |
| 7yr0mo or older | 58.8% | 59.1% |

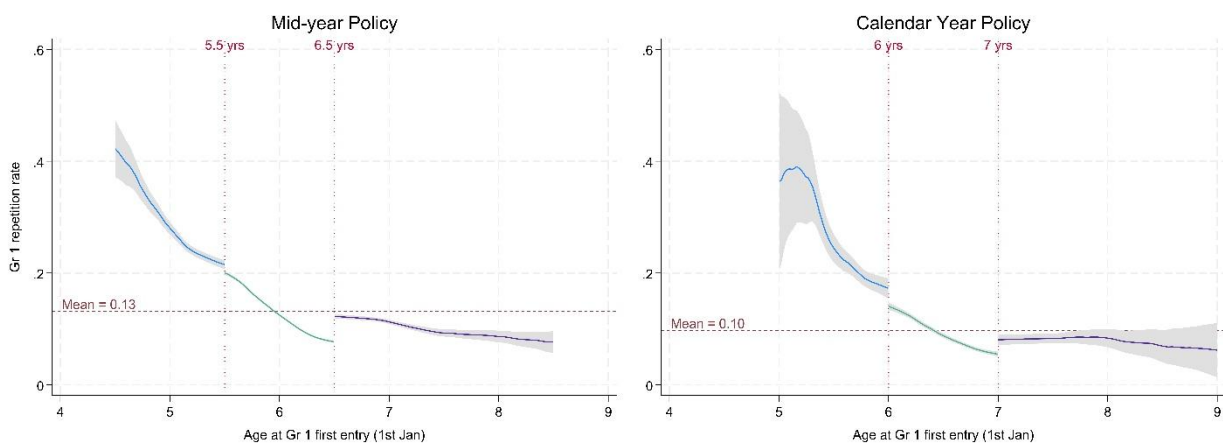
Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Only included schools that had information in the previous year, to ensure that repeaters and first-time entrants into Grade 1 could be correctly identified, provided the child stayed within the same school. Sample size = 651,566 (MY = 569,862 and CY = 81,704)

5.3 Grade 1 outcomes

For the remaining analyses, the data used comes from the balanced panel of Grade 1 first-time entrants in 2018 who were tracked through to Grade 4. Grade 1 outcomes are compared by school-entry age to assess potential relative age effects. Figure 5 plots learners' school-entry ages against the likelihood of repeating Grade 1 and shows a clear negative association: older entrants are less likely to repeat than younger peers within the same Grade 1 cohort. This pattern holds in both Mid-year and Calendar Year schools.

The Mid-year and Calendar Year age thresholds are shown as vertical dotted lines in Figure 7. Discontinuities are evident at both thresholds, with the largest discontinuity at the upper (overage) threshold for Mid-year schools. Overaged learners starting Grade 1 are substantially more likely to repeat than those just younger than them, despite starting school a year later. This provides evidence that selection into under- or overage entry is non-random.

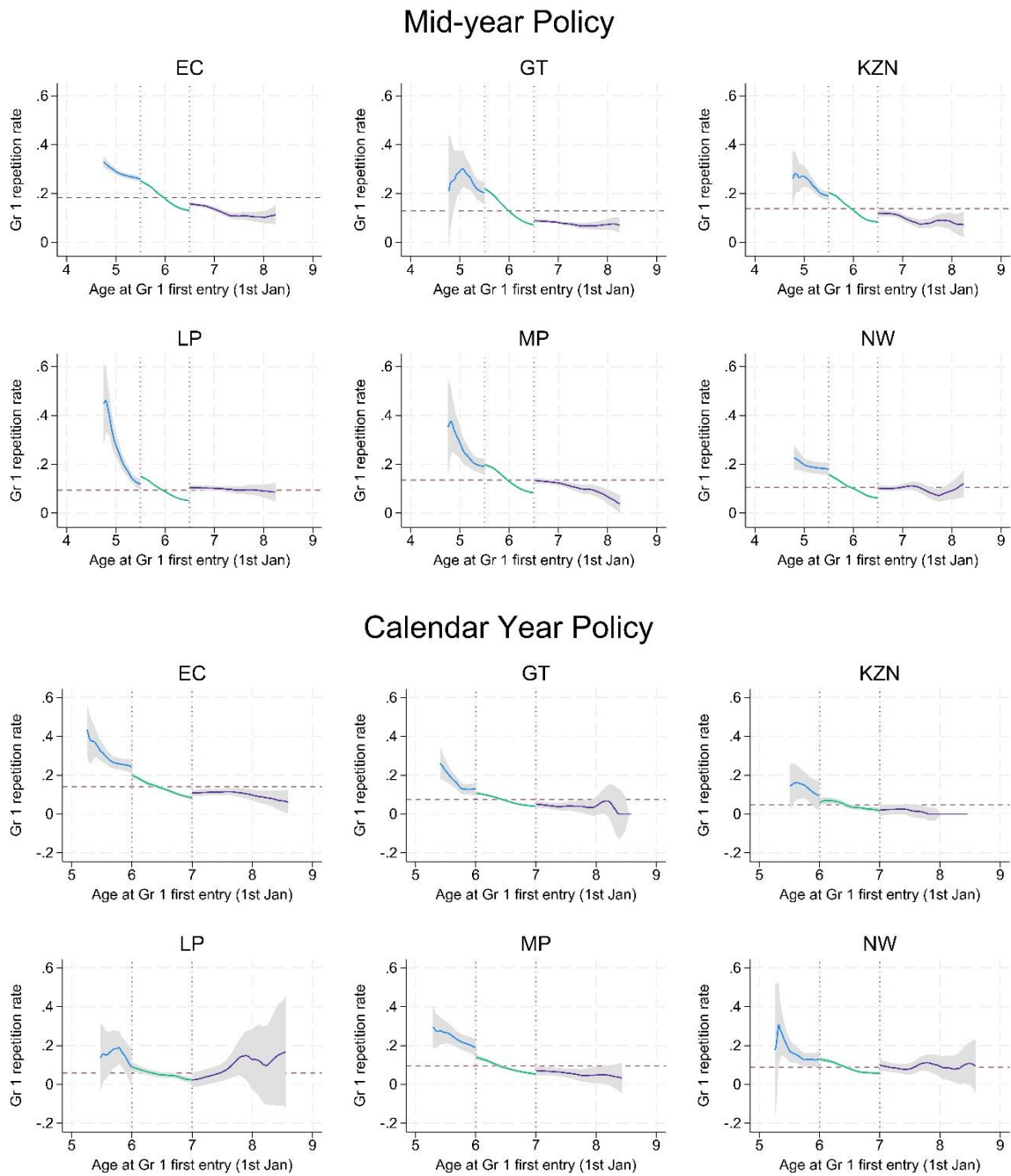
Figure 7: Grade 1 repetition rates by age of first-time school entrants in the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools



Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy. Confidence intervals at the 95% level.

This downward trend is evident across all provinces for both Mid-year and Calendar Year schools, as shown in Figure 8. However, the Grade 1 repetition rate varies substantially between provinces, ranging from 9.3% in LP to 17.6% in the EC across all schools, with Grade 1 repetition rates in Calendar Year schools lower relative to Mid-year schools. There is also corresponding variation in the gradient between school-entry age and Grade 1 repetition.

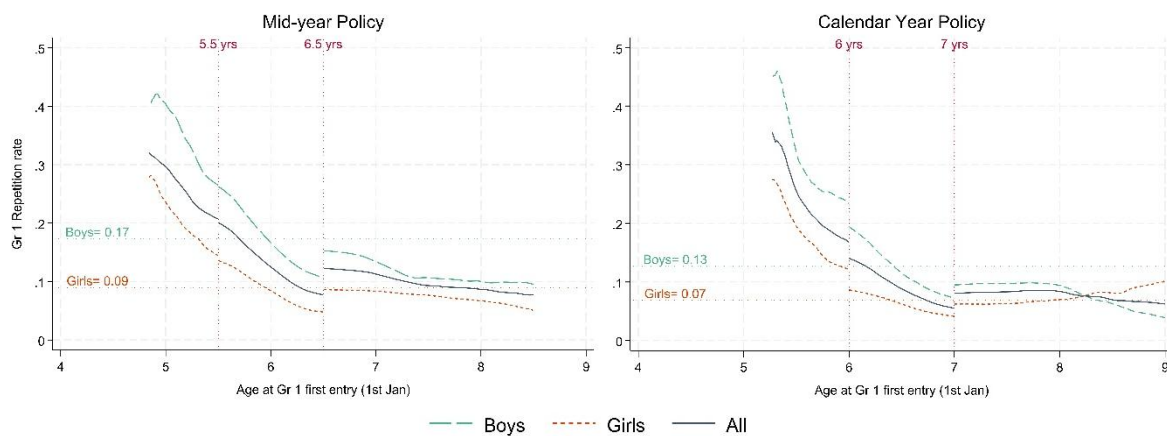
Figure 8: Grade 1 repetition rates by age of first-time school entrants in the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools, by province



Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy. Confidence intervals at the 95% level.

The differences in repetition rates by gender are significant, with girls much less likely to repeat Grade 1 than boys (9% vs 17% in Mid-year schools and 7% vs. 13% in Calendar Year schools), see Figure 9. However, given the large differences in overall repetition rates by gender, the gradients for boys and girls by school-entry age are surprisingly similar within Mid-year schools, but is noticeably steeper for boys at Calendar Year schools, suggesting the relative age effects by gender may be more pronounced at Calendar Year schools.

Figure 9: Grade 1 repetition rates by age of first-time school entrants in the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools, by gender



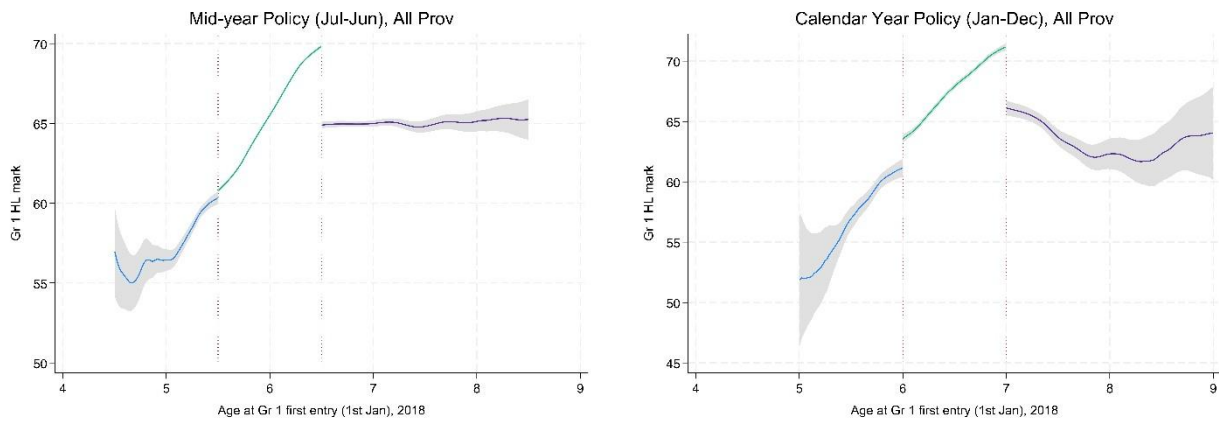
Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy.

The reverse of this pattern is seen for Grade 1 end-of-year school marks in HL, MTH and FAL. Figure 10 shows the positive relationship between age and schooling outcomes for correctly aged learners. Older learners in the cohort significantly outperform younger learners across all three subjects.

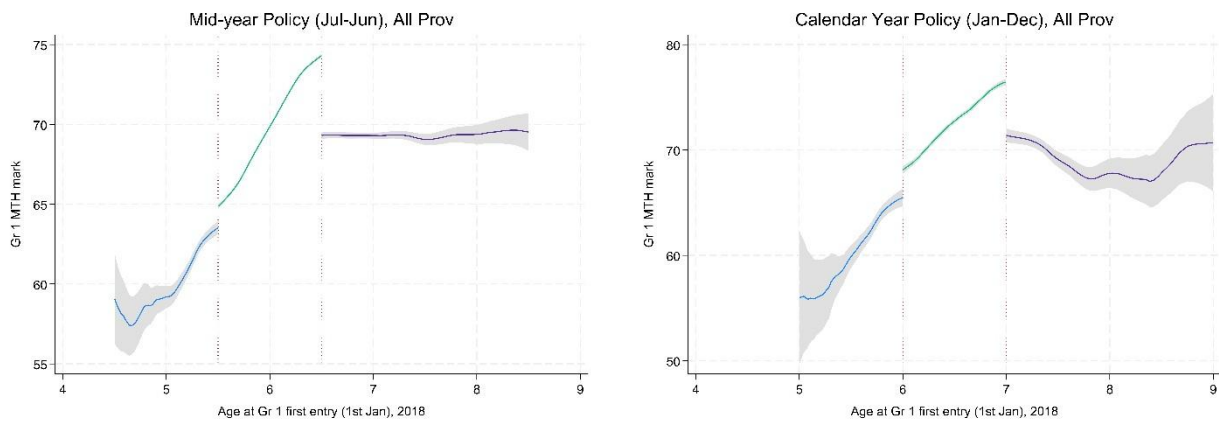
Here, too, there is a clear discontinuity at the age thresholds, but it is most pronounced at the upper age threshold, consistent with negative selection into delayed entry.

Figure 10: HL, MTH and FAL school marks by age of first-time school entrants in the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools

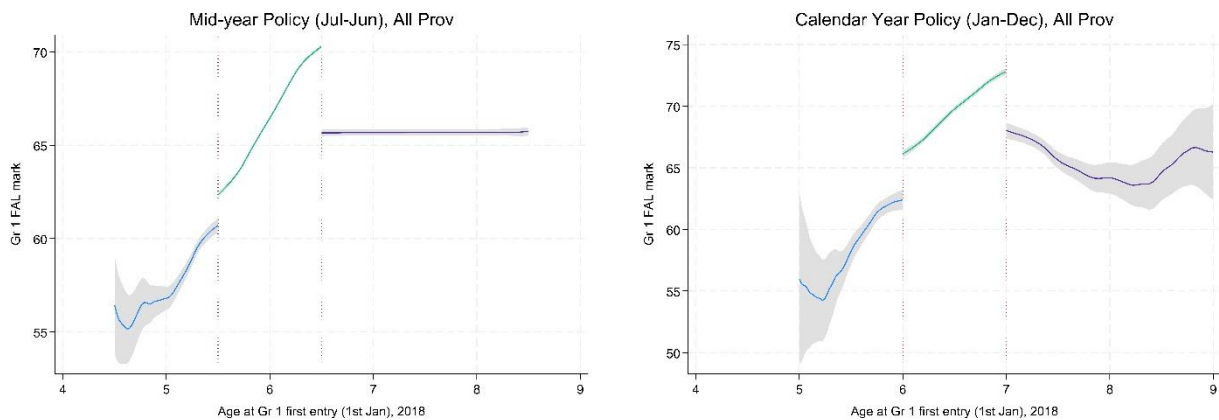
Home Language



Mathematics



First Additional Language



Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy. Confidence intervals at the 95% level.

Learner age may influence repetition decisions, as older learners tend to be less likely to repeat in higher grades. However, Clayton (2025) finds that academic performance is the primary determinant of retention, provided that learners remain eligible for repetition within that Phase³⁰. Taken together with the trends and discontinuities shown in Figure 10, this suggests that school marks, rather than age alone, are driving the observed repetition patterns.

The analysis first compares estimates across alternative functional forms. Given the endogeneity concerns discussed, ordinary least squares (OLS) estimates may not recover the causal effect and are contrasted with those from an instrumental variables approach. Table 10 and Table 11 report the estimates from both methods. The first three regressions in the tables report OLS estimates, and the subsequent three present the IV results. The Mid-year results are presented first (regressions 1-6), followed by the Calendar Year results (regressions 7-12).

Mid-year schools enrol a higher share of overage (9.6%) than underage learners (3.0%). As a result, the OLS estimates of the relative age effect are likely to be underestimated; this is evident in the results. Comparing the coefficients in Regressions (1) and (2) suggests that provincial composition may bias the first specification upward. Once school-level differences are accounted for using school fixed effects in Regression (3), the OLS estimates indicate that being one year older at entry reduces the probability of repeating Grade 1 by 10.3 percentage points. The IV estimates in regressions (4) to (6) address selection into school-entry age. They show a substantially larger effect of 15.9–16.2 percentage points. Given a mean repetition rate of 13%, this effect is large. It is also almost twice the size of the gender gap in repetition and approximately 280% as large as the gap in repetition rates between learners at Quintile 5 and Quintile 1 schools.

For Calendar Year schools, the OLS results imply that an additional year of age reduces repetition by 8.8 percentage points (Regression 9), while the IV estimates indicate a slightly larger effect of –9.8 to –10.7 percentage points (Regressions 10 to 12). Here, the difference between the OLS and IV results is less stark. Overaged learners appear to drive the bias in the relative age effects. Nevertheless, this relative age effect is also large, similar in magnitude to the mean repetition rate at Calendar Year schools and roughly 175% of the size of the gender gap³¹.

The relative age effect estimated in the IV regressions (5) and (6), and in (11) and (12), are similar. This suggests that much of the variation in repetition decisions across schools can be accounted for by province and school Quintile. Nevertheless, Regressions (6) and (12) are used as the preferred specifications, except in the percentile analyses, where the percentiles were calculated at the school level. For the percentile regressions, the functional form which includes both province and school Quintile controls (Regressions 5 and 11) is used.

30 They are eligible to repeat if they have not yet repeated a year in that Phase. See also footnote 17.

31 The gap in repetition between Quintile 5 and Quintile 1 schools is not used as a reference here because of the low number of Quintile 1 schools that follow the Calendar Year policy and the high likelihood that some of these may have been "misidentified" as Calendar Year schools due to high rates of overage admissions.

Table 10: Relative age effect on Grade 1 repetition rates for the Grade 1 2018 cohort

| Mid-year | | | | | | |
|-------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| VARIABLES | Grade 1 Repetition | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | OLS | OLS | OLS | IV | IV | IV |
| School-entry age | -0.080*** (0.001) | -0.091*** (0.001) | -0.103*** (0.001) | -0.162*** (0.002) | -0.159*** (0.002) | -0.160*** (0.002) |
| Female | | -0.086*** (0.001) | -0.087*** (0.001) | | -0.088*** (0.001) | -0.089*** (0.001) |
| Has GrR | | -0.040*** (0.001) | -0.052*** (0.001) | | -0.048*** (0.001) | -0.060*** (0.001) |
| LER | | -0.001*** (0.000) | | | -0.001*** (0.000) | |
| Constant | 0.617*** (0.007) | 0.840*** (0.008) | 0.842*** (0.007) | 1.116*** (0.014) | 1.261*** (0.015) | |
| Observations | 488 480 | 487 786 | 488 480 | 488 480 | 487 786 | 488 447 |
| R-squared | 0.011 | 0.041 | 0.150 | - | 0.034 | 0.032 |
| Quintile controls | | X | | | X | |
| Province controls | | x | | | x | |
| School FE | | | X | | | X |
| Mean of Gr 1 Rep | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 |
| Calendar Year | | | | | | |
| VARIABLES | Grade 1 Repetition | | | | | |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| | OLS | OLS | OLS | IV | IV | IV |
| School-entry age | -0.075*** (0.003) | -0.081*** (0.003) | -0.088*** (0.003) | -0.107*** (0.006) | -0.098*** (0.006) | -0.102*** (0.006) |
| Female | | -0.060*** (0.002) | -0.061*** (0.002) | | -0.061*** (0.002) | -0.061*** (0.002) |
| Has GrR | | -0.036*** (0.002) | -0.047*** (0.003) | | -0.036*** (0.002) | -0.048*** (0.003) |
| LER | | 0.002*** (0.000) | | | 0.002*** (0.000) | |
| Constant | 0.585*** (0.018) | 0.736*** (0.021) | 0.730*** (0.018) | 0.795*** (0.039) | 0.853*** (0.040) | |
| Observations | 68 070 | 67 732 | 68 070 | 68 070 | 67 732 | 68 068 |
| R-squared | 0.011 | 0.061 | 0.139 | 0.009 | 0.061 | 0.029 |
| Quintile controls | | X | | | X | |
| Province controls | | x | | | x | |
| School FE | | | X | | | X |
| Mean of Gr 1 Rep | 0.097 | 0.097 | 0.097 | 0.097 | 0.097 | 0.097 |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Gr 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Stata commands *ivregress* and *ivreghdfe* used to calculate IV estimates. Used expected school-entry age as the instrument for actual school-entry age. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 11: Relative age effect on Grade 1 Term 4 Mathematics marks for the Grade 1 2018 cohort

| Mid-year | | | | | | |
|----------------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|----------------------|
| VARIABLES | OLS | | | IV | | |
| | (1) MTH Mark | (2) MTH Mark | (3) MTH Mark | (4) MTH Mark | (5) MTH Mark | (6) MTH Mark |
| School-entry age | 5.266*** (0.059) | 6.229*** (0.057) | 7.055*** (0.053) | 13.210*** (0.127) | 13.016*** (0.120) | 13.100*** (0.107) |
| Female | | 6.506*** (0.049) | 6.624*** (0.044) | | 6.719*** (0.050) | 6.810*** (0.045) |
| Has GrR | | 3.179*** (0.060) | 4.113*** (0.062) | | 4.036*** (0.062) | 4.948*** (0.064) |
| LER | | 0.077*** (0.004) | | | 0.074*** (0.004) | |
| Constant | 37.551*** (0.356) | 15.986*** (0.389) | 20.335*** (0.332) | -10.632*** (0.769) | -26.034*** (0.759) | |
| Observations | 488 205 | 487 511 | 488 205 | 488 205 | 487 511 | 488 173 |
| R-squared | 0.016 | 0.100 | 0.301 | - | 0.074 | 0.055 |
| Quintile controls | | X | | | X | |
| Province controls | | x | | | x | |
| School FE | | | X | | | X |
| Mean of Gr1 MTH | 69.50 | 69.50 | 69.50 | 69.50 | 69.50 | 69.50 |
| Calendar Year | | | | | | |
| VARIABLES | OLS | | | IV | | |
| | (7) MTH Mark | (8) MTH Mark | (9) MTH Mark | (10) MTH Mark | (11) MTH Mark | (12) MTH Mark |
| School-entry age | 4.885*** (0.154) | 5.396*** (0.150) | 5.599*** (0.137) | 10.449*** (0.338) | 9.895*** (0.319) | 10.114*** (0.286) |
| Female | | 4.297*** (0.121) | 4.513*** (0.108) | | 4.448*** (0.122) | 4.666*** (0.109) |
| Has GrR | | 2.840*** (0.124) | 3.337*** (0.138) | | 2.881*** (0.124) | 3.409*** (0.139) |
| LER | | -0.184*** (0.012) | | | -0.170*** (0.012) | |
| Constant | 40.212*** (1.010) | 28.304*** (1.138) | 31.426*** (0.903) | 3.897* (2.207) | -1.849 (2.207) | |
| Observations | 68 004 | 67 666 | 68 004 | 68 004 | 67 666 | 68 002 |
| R-squared | 0.014 | 0.115 | 0.318 | - | 0.103 | 0.039 |
| Quintile controls | | X | | | X | |
| Province controls | | x | | | x | |
| School FE | | | X | | | X |
| Mean of Gr1 MTH | 72.10 | 72.13 | 72.10 | 72.10 | 72.13 | 72.10 |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Gr 1 and 4, from the DDD dataset 2017-2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Stata commands *ivregress* and *ivreghdfe* used to calculate IV estimates. Used expected school-entry age as the instrument for actual school-entry age. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The difference in outcomes for learners who are a year apart in age also differed significantly by school Quintile (Table 12), whilst by province, there are sizable differences in relative age effects, especially for repetition. A breakdown by province is included in the Appendix in Table 25.

The relative age effect in Mid-year schools was largest in Quintile 4 schools (17.7 percentage points), followed by Quintiles 3, 2 and 1. Independent schools had the smallest relative age effect, at only 4.6 percentage points. With Quintile 5 schools in between at 12.6 percentage points. However, this simple difference ignores the differences in mean repetition rates across the school Quintiles. Given a mean rate of only 3.3% in Mid-year independent schools, it would be impossible to have a relative age effect of 16 percentage points, as we see in Quintile 2 schools.

If instead we consider the relative age effect on repetition as a proportion of the mean repetition rate, then Quintile 5 schools have the largest gap, with the relative age effect on repetition of almost 150% of the mean repetition rate. Whilst the lowest ratio is seen in Quintile 1 schools, where the relative age effect on repetition is 114% of the mean repetition rate.

For Calendar Year schools, the largest relative age effect on Grade 1 repetition is seen in Quintile 1 and 2 schools (18.4 and 21.8 percentage points, respectively). Noting that this represents a small number of learners and a small proportion of Quintile 1 and 2 schools. The smallest effects are seen in Quintile 4 and independent schools, at just over 6 percentage points. The relative proportion of the relative age effect to the repetition rate at independent schools is a very high 170%, suggesting that very few older learners in the cohort repeat; however, this is based on the information of less than 5 000 learners.

Overall, the relative age effect appears to be largest in poorer schools. Yet once differences in average repetition rates are taken into account, the relative effect is broadly similar across socio-economic groups.

The gender differences in both the mean outcomes and relative age effects in Grade 1 are sizable, as shown in Table 13. Boys achieve lower marks and are about twice as likely to repeat Grade 1 compared to girls (in Mid-year and Calendar Year schools). Girls outperform boys in both home language and Mathematics, with a slightly larger gap in home language. The relative age effect is also larger for boys than for girls across subjects and Mid-year and Calendar Year schools. Surprisingly, the relative age effect difference between boys and girls appears to be larger in mathematics than in home language. This suggests that younger boys in Grade 1 are struggling across the board relative to their peers, but even more in numeracy than literacy.

Table 12: Relative age effect on Grade 1 outcomes, by school Quintile for the Grade 1 2018 cohort

| Mid-year | | | | | | | |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Indep. | All |
| Repetition | | | | | | | |
| Coeff | -0.158*** | -0.161*** | -0.168*** | -0.177*** | -0.126*** | -0.046*** | -0.160*** |
| SE | (0.004) | (0.004) | (0.004) | (0.008) | (0.010) | (0.011) | (0.002) |
| Mean | 0.140 | 0.129 | 0.139 | 0.136 | 0.083 | 0.033 | 0.131 |
| N | 143 296 | 133 308 | 135 403 | 42 921 | 22 698 | 10 604 | 488 447 |
| HL | | | | | | | |
| Coeff | 12.653*** | 12.772*** | 12.805*** | 13.143*** | 10.985*** | 8.472*** | 12.638*** |
| SE | (0.189) | (0.187) | (0.204) | (0.370) | (0.509) | (0.735) | (0.104) |
| Mean | 64.829 | 65.470 | 64.534 | 64.812 | 67.368 | 73.356 | 65.225 |
| N | 143 055 | 133 133 | 134 969 | 42 724 | 22 697 | 10 603 | 487 398 |
| MTH | | | | | | | |
| Coeff | 13.243*** | 13.239*** | 13.188*** | 13.422*** | 11.403*** | 8.976*** | 13.100*** |
| SE | (0.198) | (0.195) | (0.210) | (0.374) | (0.521) | (0.737) | (0.107) |
| Mean | 68.682 | 69.663 | 68.856 | 70.439 | 72.356 | 76.736 | 69.496 |
| N | 143 137 | 133 287 | 135 346 | 42 886 | 22 697 | 10 603 | 488 173 |
| FAL | | | | | | | |
| Coeff | 11.485*** | 11.415*** | 11.252*** | 11.217*** | 8.749*** | 8.128*** | 11.215*** |
| SE | (0.178) | (0.175) | (0.188) | (0.334) | (0.458) | (0.744) | (0.096) |
| Mean | 65.509 | 66.344 | 65.586 | 67.017 | 68.160 | 72.120 | 66.153 |
| N | 142 686 | 132 523 | 134 451 | 42 518 | 22 695 | 10 190 | 485 280 |

| Calendar Year | | | | | | | |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Indep. | All |
| Repetition | | | | | | | |
| Coeff | -0.184*** | -0.218*** | -0.136*** | -0.063*** | -0.084*** | -0.069*** | -0.102*** |
| SE | (0.046) | (0.033) | (0.017) | (0.012) | (0.006) | (0.019) | (0.006) |
| Mean | 0.183 | 0.158 | 0.163 | 0.081 | 0.061 | 0.037 | 0.097 |
| N | 2 290 | 4 448 | 15 603 | 11 132 | 29 361 | 4 842 | 68 068 |
| HL | | | | | | | |
| Coeff | 11.875*** | 15.566*** | 9.669*** | 9.589*** | 8.693*** | 7.022*** | 9.461*** |
| SE | (1.924) | (1.545) | (0.741) | (0.635) | (0.341) | (1.149) | (0.276) |
| Mean | 62.073 | 63.540 | 62.211 | 67.757 | 69.830 | 72.209 | 67.195 |
| N | 2 266 | 4 448 | 15 470 | 11 132 | 29 330 | 4 831 | 67 869 |
| MTH | | | | | | | |
| Coeff | 12.655*** | 14.412*** | 9.808*** | 10.327*** | 9.621*** | 8.298*** | 10.114*** |
| SE | (2.070) | (1.546) | (0.776) | (0.641) | (0.352) | (1.175) | (0.286) |
| Mean | 67.192 | 68.948 | 66.648 | 73.390 | 74.657 | 76.691 | 72.091 |
| N | 2 289 | 4 448 | 15 559 | 11 132 | 29 340 | 4 842 | 68 002 |
| FAL | | | | | | | |
| Coeff | 11.034*** | 11.202*** | 9.380*** | 8.101*** | 7.225*** | 5.197*** | 8.048*** |
| SE | (1.807) | (1.327) | (0.702) | (0.575) | (0.314) | (1.091) | (0.254) |
| Mean | 62.177 | 66.152 | 64.287 | 69.332 | 72.249 | 71.694 | 69.146 |
| N | 2 288 | 4 410 | 15 170 | 11 059 | 28 910 | 4 812 | 67 014 |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Category for learners at public schools with no school Quintile information is not shown in the table, as the sample is very small. Only 612 learners had no school Quintile information in Gr1, 219 at MY and 393 at CY schools. Used the command `ivreghdfe` in Stata to run an IV regression with expected school-entry age as an instrument for actual age of first-time Grade 1 entry with school fixed effects, controlling for gender and Grade R attendance. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 13: Relative age effect on Grade 1 outcomes, by Gender for the Grade 1 2018 cohort

| | Mid-year | | | Calendar Year | | |
|-------------------|-----------|-----------|-----------|---------------|-----------|-----------|
| | Boys | Girls | All | Boys | Girls | All |
| Repetition | | | | | | |
| Coeff | -0.203*** | -0.117*** | -0.160*** | -0.139*** | -0.064*** | -0.102*** |
| SE | (0.004) | (0.003) | (0.002) | (0.009) | (0.007) | (0.006) |
| Mean | 0.173 | 0.089 | 0.131 | 0.127 | 0.068 | 0.097 |
| N | 245 675 | 242 496 | 488 447 | 33 691 | 34 354 | 68 068 |
| HL | | | | | | |
| Coeff | 13.322*** | 12.000*** | 12.638*** | 9.877*** | 9.149*** | 9.461*** |
| SE | (0.153) | (0.141) | (0.104) | (0.412) | (0.371) | (0.276) |
| Mean | 61.791 | 68.708 | 65.225 | 64.436 | 69.904 | 67.195 |
| N | 245 148 | 241 974 | 487 398 | 33 590 | 34 256 | 67 869 |
| MTH | | | | | | |
| Coeff | 14.235*** | 11.990*** | 13.100*** | 10.702*** | 9.571*** | 10.114*** |
| SE | (0.162) | (0.143) | (0.107) | (0.429) | (0.381) | (0.286) |
| Mean | 66.345 | 72.696 | 69.496 | 69.973 | 74.173 | 72.091 |
| N | 245 536 | 242 361 | 488 173 | 33 658 | 34 321 | 68 002 |
| FAL | | | | | | |
| Coeff | 11.880*** | 10.572*** | 11.215*** | 8.256*** | 7.946*** | 8.048*** |
| SE | (0.143) | (0.131) | (0.096) | (0.378) | (0.343) | (0.254) |
| Mean | 63.263 | 69.089 | 66.153 | 67.023 | 71.227 | 69.146 |
| N | 244 062 | 240 941 | 485 280 | 33 114 | 33 877 | 67 014 |

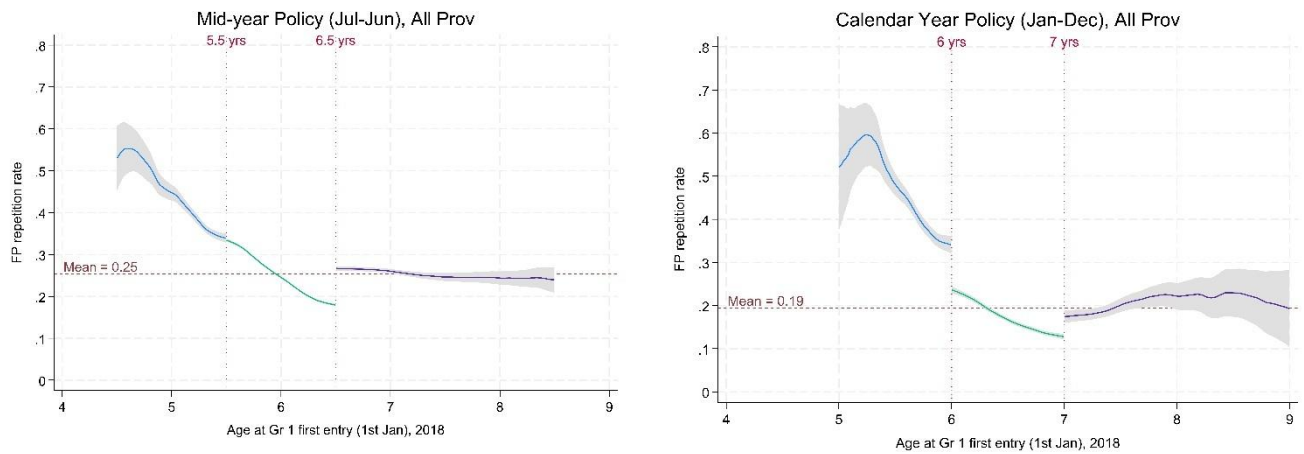
Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Used the command `ivreghdfe` in Stata to run an IV regression with expected school-entry age as an instrument for actual age of first-time Grade 1 entry with school fixed effects and controlling for Grade R attendance and (only for the total sample) for gender. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.4 Grade 4 outcomes

A similar set of analyses was conducted at the Grade 4 level to see whether the age-of-entry effects in Grade 1 persist to Grade 4. However, there are now two channels through which school-entry age affects schooling outcomes: directly and indirectly, mediated by repetition in the Foundation Phase. And, it is anticipated that these two channels would work in opposite directions. The section first presents the total effect, which is the sum of both the direct and indirect effects. The naïve approach of estimating the direct effect by controlling for the mediator would likely lead to biased results due to the same unobserved confounders as for school-entry age.

Figure 11 shows that repeating a grade at least once in the Foundation Phase is relatively common, affecting about a quarter of learners in Mid-year schools and a fifth in Calendar Year schools, and it is not age-neutral. Younger entrants are considerably more likely to repeat at least once in the Foundation Phase.

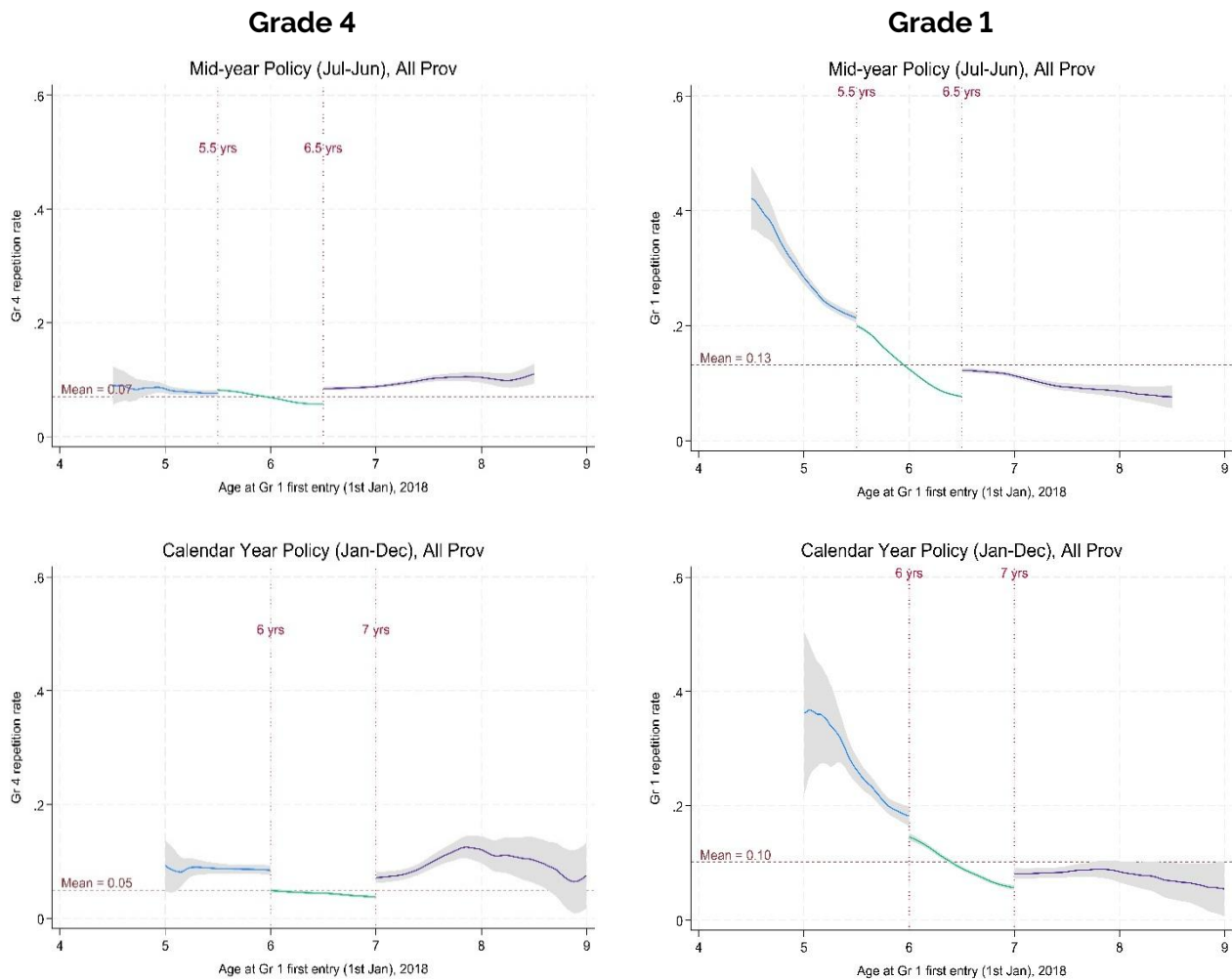
Figure 11: Foundation Phase repetition rates by age of first-time school entry for the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools



Source: DDD dataset 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW, tracking the 2018 Grade 1 first-time entrants' cohort. Sample size is 488 492 for Mid-year schools and 68 070 for Calendar Year schools. A learner is counted as having repeated in the Foundation Phase if they repeated at least once in Grades 1, 2 or 3.

Figure 12 compares Grade 4 repetition rates with those in Grade 1 by school-entry age. Repetition is far lower in Grade 4; correspondingly, the absolute gap between the youngest and oldest entrants has also narrowed. A second pattern stands out: the effect of a delayed start has not faded over time. The underlying factors that prompted parents to delay school start appear to continue shaping academic outcomes, and these may even have compounded over the years. By Grade 4, learners who entered school overage now repeat at noticeably higher rates than peers who began on time.

Figure 12: Grade 4 Repetition rates by age of first-time school entry for the 2018 Grade 1 cohort, showing Mid-year and Calendar Year schools



Source: DDD dataset 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW, tracking the 2018 Grade 1 first-time entrants' cohort. Sample size is 488 492 for Mid-year schools and 68 070 for Calendar Year schools.

Table 14 reports the two-stage least squares estimates with school fixed effects for age-of-entry effects in Grade 4 by school Quintile (a similar breakdown by province is included in the Appendix in Table 26). At Mid-year schools, the relative age effect on Foundation Phase repetition is around 4 percentage points larger than in Grade 1, but well below the overall increase in average repetition over this period (around 12 percentage points). This suggests that school-entry age mattered most for repetition in Grade 1, with weaker effects in Grades 2 and 3.

By Grade 4, a one-year difference in school-entry age still translates into a 4.2-percentage-point difference in the likelihood of repetition in Mid-year schools, where the mean repetition rate is 7.1%. The gaps in HL, MTH and FAL scores, between 5.1 and 5.9 points, or roughly 0.34 to 0.37 standard deviations, are substantive.

In Calendar Year schools in Quintiles 1 and 2, as well as independent schools, the relative age effect on repetition is insignificant. Yet differences in school marks between initially older and younger learners persist, with the exception of HL for Quintile 1 and FAL for

independent schools. Some of this may reflect higher variance as a result of smaller samples. However, repetition policies for Grades 4-9 in 2020-22 were heavily disrupted by the COVID-19 pandemic³², which could also account for the disconnect between grade repetition and academic achievement patterns.

Table 14: Relative age effect on Grade 4, by school Quintile for the Grade 1 2018 cohort

| | Mid-year | | | | | | |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Indep. | All |
| FP Rep | | | | | | | |
| Coeff | -0.212*** | -0.220*** | -0.224*** | -0.219*** | -0.180*** | -0.102*** | -0.215*** |
| SE | (0.005) | (0.005) | (0.006) | (0.009) | (0.014) | (0.019) | (0.003) |
| Mean | 0.271 | 0.253 | 0.267 | 0.236 | 0.175 | 0.106 | 0.254 |
| N | 142 861 | 133 736 | 135 915 | 43 799 | 21 689 | 9 951 | 488 288 |
| Gr 4 Rep | | | | | | | |
| Coeff | -0.043*** | -0.043*** | -0.046*** | -0.033*** | -0.025*** | -0.008 | -0.042*** |
| SE | (0.003) | (0.003) | (0.003) | (0.005) | (0.007) | (0.011) | (0.002) |
| Mean | 0.085 | 0.071 | 0.072 | 0.049 | 0.031 | 0.032 | 0.071 |
| N | 142 861 | 133 736 | 135 915 | 43 799 | 21 689 | 9 951 | 488 288 |
| HL | | | | | | | |
| Coeff | 5.258*** | 5.296*** | 4.962*** | 4.343*** | 4.935*** | 5.209*** | 5.090*** |
| SE | (0.170) | (0.168) | (0.183) | (0.313) | (0.494) | (0.723) | (0.093) |
| Mean | 64.828 | 65.955 | 65.077 | 66.503 | 67.924 | 67.783 | 65.554 |
| N | 142 660 | 133 513 | 135 617 | 43 799 | 21 689 | 9 940 | 487 555 |
| MTH | | | | | | | |
| Coeff | 5.775*** | 5.977*** | 6.023*** | 5.693*** | 5.801*** | 5.395*** | 5.886*** |
| SE | (0.176) | (0.173) | (0.191) | (0.325) | (0.545) | (0.801) | (0.097) |
| Mean | 61.703 | 63.000 | 62.306 | 63.409 | 64.747 | 64.114 | 62.562 |
| N | 142 646 | 133 625 | 135 714 | 43 778 | 21 685 | 9 951 | 487 736 |
| FAL | | | | | | | |
| Coeff | 5.482*** | 5.878*** | 5.571*** | 4.969*** | 4.827*** | 4.430*** | 5.531*** |
| SE | (0.174) | (0.174) | (0.191) | (0.317) | (0.499) | (0.770) | (0.096) |
| Mean | 59.981 | 61.049 | 60.540 | 62.624 | 64.148 | 65.586 | 60.962 |
| N | 142 612 | 133 622 | 135 774 | 43 707 | 21 639 | 9 805 | 487 495 |

32 Two circulars issued in 2021 and 2022 (National Assessment Circular No. 3 of 2021 and National Assessment Circular No. 5 of 2022) outlined changes to promotion requirements, including allowable mark adjustments of up to five percentage points and the removal of the requirement to pass Mathematics. Alongside changes to assessment practices, these measures lowered the standard required for promotion in Grade 4, reducing repetition rates. In 2020, direct "trimming" of the curriculum to focus on fundamentals as well as the cancellation of June examinations and changes to the assessment schedule (Hoadley, 2023), resulted in reduced repetition.

| | Calendar Year | | | | | | |
|-----------------|---------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Indep. | All |
| FP Rep | | | | | | | |
| Coeff | -0.147*** | -0.170*** | -0.149*** | -0.099*** | -0.097*** | -0.099*** | -0.114*** |
| SE | (0.056) | (0.039) | (0.021) | (0.016) | (0.008) | (0.030) | (0.007) |
| Mean | 0.346 | 0.285 | 0.307 | 0.148 | 0.102 | 0.092 | 0.178 |
| N | 2 269 | 4 407 | 15 372 | 11 504 | 28 853 | 4 120 | 66 955 |
| Gr 4 Rep | | | | | | | |
| Coeff | 0.014 | -0.027 | -0.023* | -0.025*** | -0.011*** | -0.003 | -0.016*** |
| SE | (0.033) | (0.024) | (0.014) | (0.009) | (0.003) | (0.017) | (0.004) |
| Mean | 0.085 | 0.080 | 0.097 | 0.037 | 0.016 | 0.024 | 0.046 |
| N | 2 269 | 4 407 | 15 372 | 11 504 | 28 853 | 4 120 | 66 955 |
| HL | | | | | | | |
| Coeff | 1.143 | 5.076*** | 4.496*** | 5.368*** | 4.569*** | 2.603** | 4.561*** |
| SE | (1.632) | (1.324) | (0.725) | (0.571) | (0.308) | (1.195) | (0.256) |
| Mean | 62.197 | 64.198 | 61.904 | 66.745 | 69.650 | 67.669 | 66.594 |
| N | 2 264 | 4 385 | 15 369 | 11 504 | 28 853 | 4 120 | 66 925 |
| MTH | | | | | | | |
| Coeff | 2.680 | 6.250*** | 4.455*** | 5.537*** | 5.136*** | 2.623* | 4.992*** |
| SE | (1.697) | (1.358) | (0.774) | (0.678) | (0.398) | (1.382) | (0.300) |
| Mean | 59.758 | 61.092 | 59.291 | 64.134 | 67.887 | 65.211 | 64.342 |
| N | 2 247 | 4 407 | 15 328 | 11 504 | 28 810 | 4 109 | 66 835 |
| FAL | | | | | | | |
| Coeff | 3.937** | 6.096*** | 5.190*** | 5.588*** | 4.812*** | 1.704 | 4.930*** |
| SE | (1.789) | (1.339) | (0.740) | (0.590) | (0.334) | (1.213) | (0.268) |
| Mean | 58.585 | 59.386 | 58.164 | 64.056 | 67.596 | 64.659 | 63.746 |
| N | 2 261 | 4 402 | 15 332 | 11 409 | 28 823 | 4 117 | 66 774 |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Category for learners at public schools with no school Quintile information is not shown in the table, as the sample is very small. Only 612 learners had no school Quintile information in Gr1, 219 at MY and 393 at CY schools. Used command *ivreghdfe* in Stata to run an IV regression with expected school-entry age as an instrument for actual age of first-time Grade 1 entry with school fixed effects and controlling for gender and Grade R attendance. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The regressions in Table 14 do not control for repetition or age-in-grade in Grade 4; repetition is therefore treated as part of the mechanism through which the relative age effect narrows over time. As a result, the estimates likely represent lower bounds of the true effect of entering school as an older learner in the cohort. Nor can the influence of an additional year of schooling simply be accounted for by adding controls for Foundation Phase repetition, as unobserved learner characteristics also act as a confounder for Foundation Phase repetition (as well as for school-entry age).

5.5 Grade 1 and 4 outcomes

The analysis then examines whether outcomes diverge between Grade 1 and Grade 4. Table 15 reports the relative age effect on learners' percentile rank within their school. In Grade 1, learners who were a year older ranked approximately 19–24 percentiles higher than their youngest peers. By Grade 4, this advantage had narrowed to 9–12 percentiles. While this convergence is encouraging, a gap of this size remains substantial, indicating that relative age effects persist into Grade 4 despite some fade-out.

Table 15: Relative age effect on Grade 1 and Grade 4 school percentiles for HL, MTH and FAL

| | Mid-year | | | Calendar Year | | |
|-----------------|------------|------------|------------|---------------|------------|------------|
| | Gr 1 Perc. | Gr 4 Perc. | Gr 4 Perc. | Gr 1 Perc. | Gr 4 Perc. | Gr 4 Perc. |
| HL | | | | | | |
| Coeff | 0.242*** | 0.114*** | -0.018*** | 0.199*** | 0.103*** | -0.021*** |
| SE | (0.002) | (0.002) | (0.002) | (0.006) | (0.006) | (0.005) |
| N | 486 738 | 487 054 | 486 008 | 67 533 | 67 694 | 67 495 |
| R ² | 0.062 | 0.111 | 0.394 | 0.045 | 0.059 | 0.438 |
| MTH | | | | | | |
| Coeff | 0.244*** | 0.124*** | -0.012*** | 0.214*** | 0.094*** | -0.042*** |
| SE | (0.002) | (0.002) | (0.002) | (0.006) | (0.006) | (0.005) |
| N | 487 511 | 487 235 | 486 963 | 67 666 | 67 610 | 67 544 |
| R ² | 0.050 | 0.044 | 0.343 | 0.026 | 0.003 | 0.393 |
| FAL | | | | | | |
| Coeff | 0.231*** | 0.113*** | -0.003** | 0.188*** | 0.106*** | 0.008* |
| SE | (0.002) | (0.002) | (0.002) | (0.006) | (0.006) | (0.005) |
| N | 484 618 | 486 997 | 483 832 | 66 705 | 67 550 | 66 523 |
| R ² | 0.049 | 0.103 | 0.352 | 0.032 | 0.059 | 0.333 |
| Controls | | | | | | |
| Standard | x | x | x | x | x | x |
| Gr1 Perc | | | x | | | x |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Used the command *ivregress* in Stata to run an IV regression with expected school-entry age as an instrument for actual age of first-time Grade 1 entry. The Standard Controls were included in all of the Regressions, these included Gender, Has Grade R, School Learner Educator Ratio, school province and school Quintile dummies. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To identify which groups of learners improve their relative position over time, the analysis employs Staying Probabilities and Directional Rank Probabilities, reported in Table 16 to Table 18. Learners are grouped by birth quartile, defined relative to each school's entry-age policy: in Mid-year schools, Quartile 1 covers July–September births and Quartile 2 covers October–December, whereas in Calendar Year schools, Quartile 1 spans January–March. Learners are also classified into subject mark quartiles based on their within-school Grade 1 performance in each subject. This approach, therefore, does not use each learner's actual school-entry age, but instead proxies likely school-entry age using birth quartile.

Across all subject mark quartiles, younger birth-quartile learners were slightly less likely than their older peers to remain in the same performance quartile by Grade 4. This pattern is mainly evident when comparing quartiles 1 and 4 and is otherwise fairly small. While downward mobility from the lowest quartiles is expected, it is less obvious why younger high-performing learners would be more likely to shift out of the top quartile. The trend, however, is consistent across both Mid-year and Calendar Year schools.

Mobility from the middle quartiles is, predictably, higher, but Staying Probabilities do not reveal whether this movement is upward or downward. Upward Rank Probabilities in Table 17 offer clearer insight. The largest gaps are seen among learners in the lowest subject quartiles. In Mid-year schools, Grade 1 learners in the lowest HL quartile who were born in Quartile 4 were five percentage points more likely to move up by two percentiles or more

than those born in Quartile 1. Among top-quartile HL learners, those from Quartile 4 were two percentage points more likely to rise than their Quartile 1 peers.

The inverse is tested in Table 18, showing the Downward Rank probabilities. Here, the likelihood of moving down the percentile ranking was lower for younger birth-quartile learners. This difference, especially in MTH and FAL, was more evident for learners in the lower half of academic achievement in grade 1. Overall, learners who would have been younger at school entry show a consistently higher probability of moving up the rankings and a lower probability of moving down than their older classmates.

The main analysis uses an upward or downward movement (delta) of two percentiles. Results for alternative deltas of zero and five percentiles are reported in the Appendix (Table 27 to Table 30) and reveal similar patterns, suggesting that the findings are robust and apply continuously across the distribution.

Table 16: Staying Probabilities between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| | | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|--------------|----------------|-----|-----|--------------|
| Subject | Gr 1 Subject Quartile | Birth Quartile | | | | Birth Quartile | | | |
| | | 1 (Oldest) | 2 | 3 | 4 (Youngest) | 1 (Oldest) | 2 | 3 | 4 (Youngest) |
| HL | 1 | 53% | 53% | 51% | 50% | 55% | 56% | 55% | 54% |
| HL | 2 | 33% | 33% | 33% | 33% | 35% | 36% | 35% | 35% |
| HL | 3 | 32% | 33% | 34% | 34% | 35% | 36% | 36% | 36% |
| HL | 4 | 59% | 57% | 57% | 57% | 62% | 61% | 62% | 60% |
| MTH | 1 | 53% | 53% | 51% | 50% | 56% | 55% | 55% | 53% |
| MTH | 2 | 33% | 33% | 33% | 33% | 35% | 35% | 34% | 34% |
| MTH | 3 | 32% | 32% | 32% | 32% | 35% | 34% | 34% | 35% |
| MTH | 4 | 58% | 56% | 55% | 54% | 59% | 60% | 57% | 60% |
| FAL | 1 | 51% | 51% | 50% | 50% | 49% | 49% | 50% | 50% |
| FAL | 2 | 32% | 32% | 32% | 32% | 33% | 31% | 32% | 31% |
| FAL | 3 | 31% | 32% | 32% | 32% | 32% | 32% | 31% | 33% |
| FAL | 4 | 56% | 55% | 54% | 53% | 57% | 55% | 54% | 53% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort.

Table 17: Upward Rank Probabilities ($\delta= 0.02$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| | | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|--------------|----------------|-----|-----|--------------|
| Subject | Gr 1 Subject Quartile | Birth Quartile | | | | Birth Quartile | | | |
| | | 1 (Oldest) | 2 | 3 | 4 (Youngest) | 1 (Oldest) | 2 | 3 | 4 (Youngest) |
| HL | 1 | 64% | 66% | 67% | 69% | 63% | 63% | 64% | 66% |
| HL | 2 | 47% | 49% | 51% | 52% | 47% | 47% | 48% | 53% |
| HL | 3 | 37% | 39% | 40% | 40% | 39% | 40% | 41% | 42% |
| HL | 4 | 25% | 25% | 26% | 27% | 26% | 28% | 29% | 29% |
| MTH | 1 | 65% | 65% | 68% | 70% | 64% | 64% | 64% | 68% |
| MTH | 2 | 48% | 50% | 51% | 52% | 48% | 51% | 51% | 54% |
| MTH | 3 | 38% | 38% | 40% | 40% | 39% | 41% | 42% | 44% |
| MTH | 4 | 25% | 25% | 26% | 26% | 26% | 27% | 26% | 29% |
| FAL | 1 | 66% | 67% | 69% | 70% | 68% | 69% | 69% | 70% |
| FAL | 2 | 50% | 51% | 53% | 53% | 54% | 54% | 54% | 54% |
| FAL | 3 | 39% | 40% | 40% | 40% | 41% | 42% | 43% | 43% |
| FAL | 4 | 25% | 25% | 25% | 25% | 25% | 25% | 26% | 26% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in [0; 1 - \delta)$ are included in the analysis.

Table 18: Downward Rank Probabilities ($\delta= 0.02$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| | | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|--------------|----------------|-----|-----|--------------|
| Subject | Gr 1 Subject Quartile | Birth Quartile | | | | Birth Quartile | | | |
| | | 1 (Oldest) | 2 | 3 | 4 (Youngest) | 1 (Oldest) | 2 | 3 | 4 (Youngest) |
| HL | 1 | 26% | 25% | 23% | 22% | 27% | 27% | 26% | 23% |
| HL | 2 | 48% | 46% | 44% | 42% | 47% | 46% | 46% | 41% |
| HL | 3 | 57% | 55% | 55% | 54% | 55% | 54% | 53% | 52% |
| HL | 4 | 66% | 65% | 65% | 64% | 63% | 62% | 61% | 61% |
| MTH | 1 | 25% | 25% | 23% | 21% | 27% | 26% | 26% | 22% |
| MTH | 2 | 46% | 45% | 43% | 42% | 45% | 43% | 43% | 40% |
| MTH | 3 | 56% | 56% | 54% | 55% | 55% | 53% | 52% | 50% |
| MTH | 4 | 66% | 66% | 65% | 66% | 65% | 63% | 64% | 61% |
| FAL | 1 | 25% | 23% | 22% | 20% | 23% | 23% | 22% | 20% |
| FAL | 2 | 44% | 44% | 42% | 42% | 40% | 41% | 40% | 40% |
| FAL | 3 | 56% | 55% | 54% | 54% | 53% | 53% | 51% | 51% |
| FAL | 4 | 67% | 66% | 66% | 67% | 66% | 66% | 66% | 66% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in (0 + \delta; 1]$ are included in the analysis.

6. Discussion and recommendations

This study documents meaningful differences in how admissions-age policy is applied across South African schools. Two broad policies operate in practice: a Mid-year policy, typically followed by poorer and more rural schools, and a Calendar Year policy, more common in wealthier or urban schools. The former generally admits children at younger ages, while in the latter, they are enrolled 6 months later on average.

The findings indicate that the youngest learners within a cohort achieve the lowest end-of-year marks and face the highest risk of grade repetition during the Foundation Phase. These patterns reflect relative age effects, where younger learners perform worse than their older classmates within the same grade. One limitation of the analysis is that school marks are norm-referenced and assessments are set at the school level, meaning that marks are not standardised across schools. However, this concern is largely mitigated by the inclusion of school fixed effects. In addition, the results could be affected by teacher marking bias if teachers systematically marked either younger or older learners more harshly than others in the same class. If such a bias exists, the direction is unclear. It would need to be ascertained by comparing school marks against standardised test results.

In principle, comparing the outcomes across Mid-year and Calendar Year schools could allow for the calculation of the effects of differences in absolute school-entry age (studies that have done this include Cascio & Schanzenbach (2016) and Görlitz et al. (2022)). However, the administrative data currently available do not permit a clean comparison across the two policy regimes. School marks are not standardised, making them an unsuitable measure for comparing outcomes across the two policies. Without standardised assessment results, there is currently no robust empirical basis to advocate for a shift in either direction.

Even if conclusive evidence existed regarding which school-entry age range better aligns the average learner's cognitive and socio-emotional development with curriculum demands, policy decisions should not be made on this basis alone. Any assessment of school readiness must consider learners' realistic alternatives. While a learner entering Grade 1 at age five and a half may be less cognitively prepared than they would be six months or a year later, this cannot be evaluated in isolation. The stimulation and nutrition provided by public schools must be assessed against what learners would otherwise receive at home or in early learning centres. For children from poorer households, boys in particular, the gains they achieve at school, from both formal instruction and their peers, may outweigh the risks and costs of repetition (Elder & Lubotsky, 2009).

Nonetheless, the coexistence of two school-entry age norms within an already unequal system may contribute to the deepening of inequalities at the macro-level. Learners who are least school-ready are disproportionately enrolled in schools where earlier entry is typical. Mid-year schools therefore, face the compounded challenge of serving learners who are both less developmentally prepared and relatively younger and less mature. This increases within-class heterogeneity in schools serving more disadvantaged communities. However, these patterns may reflect constrained choices rather than suboptimal decision-making: for many children from less well-resourced households, entering school at the earliest possible age may represent the best available option.

Recent policy efforts to expand access to, and improve the financing and quality of, early learning and care, alongside the move to make Grade R compulsory, are likely to increase the range of early learning options available to families before Grade 1 entry. In light of these changes, a review of the school-entry age policy may be warranted in the next few years. Under improved early learning conditions, allowing some learners currently attending Mid-year schools to enter Grade 1 at age 6 rather than 5 may be preferable, particularly for children born in the first half of the year. However, this conclusion is highly contingent on the quality of provision. Poor-quality programmes are unlikely to yield learning gains (Samuels et al., 2015). Under such conditions, entering Grade 1 earlier may remain the more effective option for learning, despite the risk and costs of repetition.

In the absence of clear evidence favouring one policy regime, several practical policy responses remain open. Firstly, continuing to improve the quality of instruction at the Grade 1-3 level would benefit all learners and may mitigate some of the relative age effects, especially if repetition rates are reduced.

In the absence of clear evidence favouring one policy regime, several practical policy responses remain available. First, continuing to improve the quality of instruction in Grades 1-3 would benefit all learners and may mitigate some of the relative age effects, particularly if it also reduces repetition rates.

A second step is to raise awareness of the implications of age, maturity and school readiness among teachers, ECD practitioners, officials and parents. Better information could gradually shift local norms toward the Calendar Year policy for Grade R entry at schools where such a shift is desirable. The existing legislation already permits this; no legislative change is necessary. Any such shift would need to be carefully managed to avoid widening the age range within Grade 1, ideally keeping it close to one year.

Potential adjustments must also account for existing regulatory constraints. Schools and School Governing Bodies (SGBs) are prohibited from conducting school readiness testing for school admission (South African Schools Act, 1996), limiting direct assessment of children's developmental preparedness. This places greater weight on parental decisions. Strengthening guidance to parents and early learning practitioners, and improving the consistency of information flows, may therefore be among the most practical short-term levers for reducing early-grade inequities associated with age and readiness at entry.

This research also only extended to Grade 4, where evidence of a relative age effect was still evident, despite some fade-out. This raises the question of whether this effect persists into later school grades, the tertiary system and the labour market.

Lastly, this research raises broader questions about the impact that school-entry age policy choices have on the effectiveness of grade repetition and the absolute levels of learning achieved by certain grade levels. For example, could a shift in age make a discernible difference in literacy and numeracy development? Addressing these questions will require further empirical research. A recent nationally representative early-grade reading assessment³³ covering Grade 1-4 could provide an appropriate basis for evaluating

33 The Funda Uphumelele National Survey was run in 2025 by the Department of Basic Education.

the effects of school-entry age policy on learning outcomes. The release of this data and future waves would make such analysis increasingly feasible.

7. Conclusion

This paper provides the first systematic analysis of school-entry age and academic outcomes across much of South Africa. It contributes to a growing international literature on relative age effects in low- and middle-income countries and documents how these effects operate in a context where early-grade repetition remains common. The findings show that school-entry age has a substantial influence on grade progression during the Foundation Phase. The youngest learners within a cohort consistently achieve lower marks and face the highest risk of repetition, whilst the relative age effect is particularly pronounced for boys and learners attending poorer schools.

Encouragingly, compliance with the 18-month school-entry age window set out in the South African Schools Act (Act 84 of 1996) is high across the six provinces examined and has improved since 2018. Most schools follow a dominant entry-age norm, either Mid-year or Calendar Year, which keeps the age range within cohorts to roughly one year and helps reduce initial heterogeneity in classrooms. Underage entry has continued to decline, which is positive given the weaker academic outcomes and higher repetition rates typically observed among underage learners.

Further reductions in underage entry into both Grade R and Grade 1 may therefore be beneficial, as high repetition rates impose additional costs on the system and may undermine learners' confidence. Strengthening engagement with parents remains important, particularly for children born in the first half of the year, who may benefit from starting Grade R a year later or spending an additional year in Grade R before entering Grade 1, and this is already permissible under the national policy. At the same time, overage entry should not be too strongly discouraged where it reflects genuine developmental needs.

The results also highlight broader concerns about gender disparities in educational outcomes (Wills, 2025). Growing attention has been paid to the pro-girl gap in South African schooling, and these findings provide further evidence of boys' relative underperformance. The relative age gap due to school-entry age narrows by Grade 4, partly because younger learners who repeat age by one year, and gain an additional year of instruction (Clayton, 2025). However, a relative age gap remains evident, particularly in Mid-year schools.

The findings underscore the close link between age of entry and school readiness. As South Africa moves toward universal Grade R, aligning school-entry practices more closely with children's developmental preparedness will be important for reducing early-grade inequalities. Future research could further examine differences in outcomes between the two school-entry regimes, particularly if standardised learning assessments become more widely available, to better understand the medium- and long-term consequences of earlier or later school entry.

8. References

- Altwicker-Hámori, S., & Köllő, J. (2012). Whose children gain from starting school later? – Evidence from Hungary. *Educational Research and Evaluation*, 18(5), 459–488. <https://doi.org/10.1080/13803611.2012.695142>
- Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014. <https://doi.org/10.2307/2937954>
- Angrist, J. D., & Krueger, A. B. (1992). The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association*, 87(418), 328–336.
- Arroyo, V., Díaz, J., Carmona, R., Ortiz, C., & Linares, C. (2016). Impact of air pollution and temperature on adverse birth outcomes: Madrid, 2001–2009. *Environmental Pollution*, 218, 1154–1161. <https://doi.org/10.1016/j.envpol.2016.08.069>
- Ballatore, R. M., Paccagnella, M., & Tonello, M. (2020). Bullied because younger than my mates? The effect of age rank on victimisation at school. *Labour Economics*, 62, 101772. <https://doi.org/10.1016/j.labeco.2019.101772>
- Barua, R., & Lang, K. (2016). School Entry, Educational Attainment, and Quarter of Birth: A Cautionary Tale of a Local Average Treatment Effect. *Journal of Human Capital*, 10(3), 347–376.
- Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Quarterly Journal of Economics*, 121(4), 1437–1472. (24091802). <https://doi.org/10.1162/qjec.121.4.1437>
- Bhattacharya, D., & Mazumder, B. (2011). A nonparametric analysis of black-white differences in intergenerational income mobility in the United States: Intergenerational income mobility. *Quantitative Economics*, 2(3), 335–379. <https://doi.org/10.3982/QE69>
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2), 455–467.
- Bound, J., & Jaeger, D. A. (1996). *On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist & Krueger's "Does Compulsory School Attendance Affect Schooling and Earnings?"* (Working Paper No. 5835; NBER Working Paper Series). National Bureau of Economic Research.
- Bound, J., & Jaeger, D. A. (2000). Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? In *Research in labor economics* (Vol. 19, pp. 83–108). Emerald Group Publishing Limited. <https://www.djaeger.org/research/pubs/rle2000.pdf>
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430), 443–450. <https://doi.org/10.2307/2291055>
- Buckles, K. S., & Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *The Review of Economics and Statistics*, 95(3), 711–724.
- Cascio, E. U., & Schanzenbach, D. W. (2016). First in the Class? Age and the Education Production Function. *Education Finance and Policy*, 11(3), 225–250.

- Clayton, R. (2025). *The impact of early grade repetition on test scores: Evidence from a regression discontinuity design in South Africa* (Working Paper No. 06/25). Research on Socioeconomic Policy.
- Corman, H. (2003). The effects of state policies, individual characteristics, family characteristics, and neighbourhood characteristics on grade repetition in the United States. *Economics of Education Review*, 22(4), 409–420. [https://doi.org/10.1016/S0272-7757\(02\)00070-5](https://doi.org/10.1016/S0272-7757(02)00070-5)
- Crawford, C., Dearden, L., & Greaves, E. (2014). The Drivers of Month-of-Birth Differences in Children's Cognitive and Non-Cognitive Skills. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 177(4), 829–860. <https://doi.org/10.1111/rssa.12071>
- Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, 25(1), 43–62. <https://doi.org/10.1016/j.econedurev.2004.10.004>
- Department of Basic Education. (2023). *Grade promotion, repetition and dropping out 2018 to 2021* (Data Report). Department of Basic Education. <https://www.education.gov.za/LinkClick.aspx?fileticket=llofwGUQtBo%3d&tabid=92&portalid=0&mid=4362&forcedownload=true>
- Dhuey, E., & Lipscomb, S. (2008). What makes a leader? Relative age and high school leadership. *Economics of Education Review*, 27(2), 173–183. <https://doi.org/10.1016/j.econedurev.2006.08.005>
- Diris, R. (2017). Don't Hold Back? The Effect of Grade Retention on Student Achievement. *Education Finance and Policy*, 12(3), 312–341. https://doi.org/10.1162/EDFP_a_00203
- Dobkin, C., & Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review*, 29(1), 40–54. <https://doi.org/10.1016/j.econedurev.2009.04.003>
- Dorélien, A. M. (2016). Birth seasonality in Sub-Saharan Africa. *Demographic Research*, 34, 761–796.
- Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers. *The Journal of Human Resources*, 44(3), 641–683.
- Fredriksson, P., & Öckert, B. (2005). Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.760728>
- Givord, P. (2020). *How student's month of birth is linked to performance at school: New evidence from PISA* (OECD Education Working Papers No. 221; OECD Education Working Papers, Vol. 221). <https://doi.org/10.1787/822ea6ce-en>
- Görlitz, K., Penny, M., & Tamm, M. (2022). The long-term effect of age at school entry on cognitive competencies in adulthood. *Journal of Economic Behavior & Organization*, 194, 91–104. <https://doi.org/10.1016/j.jebo.2021.12.009>
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251–1271. <https://doi.org/10.2307/1913827>
- Herbst, M., & Strawiński, P. (2016). Early effects of an early start: Evidence from lowering the school starting age in Poland. *Journal of Policy Modeling*, 38(2), 256–271. <https://doi.org/10.1016/j.jpolmod.2016.01.004>

- Hoadley, U. (2023). *COVID-19 and the South African curriculum policy response* (COVID Generation) [Research Note]. RESEP. https://resep.sun.ac.za/wp-content/uploads/2023/05/Hoadley_2023_COVID-19_curriculum_policy_response_SA_15May2023.pdf
- Huang, F. L., & Invernizzi, M. A. (2013). Birthday effects and preschool attendance. *Early Childhood Research Quarterly*, 28(1), 11–23. <https://doi.org/10.1016/j.ecresq.2012.03.002>
- Jerrim, J., Lopez-Agudo, L. A., & Marcenaro-Gutierrez, O. D. (2022). Grade retention and school entry age in Spain: A structural problem. *Educational Assessment, Evaluation and Accountability*, 34(3), 331–359. <https://doi.org/10.1007/s11092-021-09375-7>
- Kawaguchi, D. (2011). Actual age at school entry, educational outcomes, and earnings. *Journal of the Japanese and International Economies*, 25(2), 64–80. <https://doi.org/10.1016/j.jjie.2009.02.002>
- McDonough, I. K. (2015). Dynamics of the black–white gap in academic achievement. *Economics of Education Review*, 47, 17–33. <https://doi.org/10.1016/j.econedurev.2015.03.007>
- McEwan, P. J., & Shapiro, J. S. (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *The Journal of Human Resources*, 43(1), 1–29.
- Morrison, F. J., Alberts, D. M., & Griffith, E. M. (1997). Nature–nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, 33(2), 254–262. <https://doi.org/10.1037/0012-1649.33.2.254>
- Mühlenweg, A. M. (2010). Young and innocent: International evidence on age effects within grades on victimization in elementary school. *Economics Letters*, 109(3), 157–160. <https://doi.org/10.1016/j.econlet.2010.08.032>
- Nam, K. (2014). Until when does the effect of age on academic achievement persist? Evidence from Korean data. *Economics of Education Review*, 40, 106–122. <https://doi.org/10.1016/j.econedurev.2014.02.002>
- Peck, B., & Trimmer, K. (1995). The late birthday effect in Western Australia. *Issues in Educational Research*, 5(1), 35–52.
- Pedraja-Chaparro, F., Santín, D., & Simancas, R. (2015). Determinants of grade retention in France and Spain: Does birth month matter? *Journal of Policy Modeling*, 37(5), 820–834. <https://doi.org/10.1016/j.jpolmod.2015.04.004>
- Pehkonen, J., Viinikainen, J., Böckerman, P., Pulkki-Råback, L., Keltikangas-Järvinen, L., & Raitakari, O. (2015). Relative age at school entry, school performance and long-term labour market outcomes. *Applied Economics Letters*, 22(16), 1345–1348. <https://doi.org/10.1080/13504851.2015.1031864>
- Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56, 152–176. <https://doi.org/10.1016/j.econedurev.2016.12.001>
- Platas, L. M., Perry, L., Piper, B., Sitabkhan, Y., & Ketterlin-Geller, L. (2022). School-entry predictors of lower primary reading and mathematics achievement in Kenya. *Research in Comparative and International Education*, 17(3), 441–459. <https://doi.org/10.1177/17454999221084414>
- Proctor, T. B., Black, K. N., & Feldhusen, J. F. (1986). Early Admission of Selected Children to Elementary School: A Review of the Research Literature. *The Journal of Educational Research*, 80(2), 70–76.

- Puhani, P. A., & Weber, A. M. (2007). Does the early bird catch the worm?: Instrumental variable estimates of early educational effects of age of school entry in Germany. *Empirical Economics*, 32(2–3), 359. <https://doi.org/10.1007/s00181-006-0089-y>
- Robertson, E. (2011). The effects of quarter of birth on academic outcomes at the elementary school level. *Economics of Education Review*, 30(2), 300–311. <https://doi.org/10.1016/j.econedurev.2010.10.005>
- Samuels, M.-L., Taylor, S., Shepherd, D., Van Der Berg, S., Jacob, C., Deliwe, C. N., & Mabogoane, T. (2015). Reflecting on an impact evaluation of the Grade R programme: Method, results and policy responses. *African Evaluation Journal*, 3(1), 10 pages. <https://doi.org/10.4102/aej.v3i1.139>
- Sandercock, G., Ogunleye, A., Parry, D., Cohen, D., Taylor, M., & Voss, C. (2014). Athletic Performance and Birth Month: Is the Relative Age Effect More than just Selection Bias? *International Journal of Sports Medicine*, 35(12), 1017–1023. <https://doi.org/10.1055/s-0034-1368725>
- Shepard, L. A., Taylor, G. A., & Kagan, S. L. (1996). *Trends in Early Childhood Assessment Policies and Practices* (ED 450926). Office of Educational Research and Improvement.
- Solli, I. F. (2017). Left behind by birth month. *Education Economics*, 25(4), 323–346. <https://doi.org/10.1080/09645292.2017.1287881>
- South African Schools Act, Act 84, Republic of South Africa, 377 17579 (1996).
- Spaull, N., & Jansen, J. D. (Eds.). (2019). *South African schooling: The enigma of inequality: a study of the present situation and future possibilities* (Vol. 10). Springer.
- Strand, L. B., Barnett, A. G., & Tong, S. (2011). The influence of season and ambient temperature on birth outcomes: A review of the epidemiological literature. *Environmental Research*, 111(3), 451–462. <https://doi.org/10.1016/j.envres.2011.01.023>
- Symul, L., Hsieh, P., Shea, A., Moreno, Crc., Skene, D. J., Holmes, S., & Martinez, M. (2022). *Unmasking Seasonal Cycles in Human Fertility: How holiday sex and fertility cycles shape birth seasonality*. *Sexual and Reproductive Health*. <https://doi.org/10.1101/2020.11.19.20235010>
- Valdés, M. T. (2024). The effect of the month of birth on academic achievement: Heterogeneity by social origin and gender. *European Societies*, 26(4), 1094–1120. <https://doi.org/10.1080/14616696.2023.2289652>
- Western Cape Department of Education. (2025, November 12). *Learner Admissions FAQ | Western Cape Government*. <https://www.westerncape.gov.za/education/learner-admissions-faq>
- Wills, G. (2025, August 26). *Gender Patterns In South African Education*. Quantitative Education Research Conference.
- Wilson, G. (1999). The Birthdate Effect in School Sports Teams. *European Journal of Physical Education*, 4(2), 139–145. <https://doi.org/10.1080/1740898990040203>
- Yuret, T. (2018). Birthday effect. *Applied Economics Letters*, 25(13), 914–917. <https://doi.org/10.1080/13504851.2017.1383592>
- Zweimüller, M. (2013). The Effects of School Entry Laws on Educational Attainment and Starting Wages in an Early Tracking System. *Annals of Economics and Statistics*, (111/112), 141–169. <https://doi.org/10.2307/23646329>

9. Appendix

9.1 History of the School Admission's Age Policy

There are two different sections of education policy in the South African Schools Act (No 84 of 1996), which govern the age at which children enter school in South Africa. The first of these relates to the minimum age of admission, whilst the second outlines the compulsory school ages. A list of the relevant legislation since 1996 is provided in Table 19.

Table 19: Legislation governing the age of admission to Grade R and Grade 1 in South Africa

| Legislation | Date of Implementation | Grade 1 age of admission |
|---|------------------------|---|
| South African Schools Act No. 84 of 1996 | 1 January 1997 | Not specified |
| GN 19377, 19 October 1998 | 1 January 2000 | Year in which turn 7 |
| GN 1356, 11 December 2001 | 1 January 2002 | Allowed deviations at discretion of Head of Department |
| Act No. 50 of 2002 | 1 January 2004 | 5 turning 6 by 30 June (4 turning 5 by 30 June for Gr R) |
| Basic Education Laws Amendment Act 32 of 2024 | Not specified | Not specified for Gr 1. Instead start Gr R in year which turn 6 |

Source: South African Law. Table builds on work presented by Cally Ardington and Alicia Menendez

9.1.1 Minimum Admissions Age

In the first South African Schools Act No 84 of 1996 (No. 17579), no explicit ages were codified. However, in Section 5 – Admission to Public Schools, the following is noted:

“(4) The Minister may by notice in the Government Gazette, after consultation with the Council of Education Ministers, determine age requirements for the admission of learners to a school or different grades at a school.”

This was clarified 2 years later in the 1998 Government Gazette No.19377 (Notice 2433 of 1998) which clarified the Admissions age as follows:

“3. The statistical age norm per grade is the grade number plus 6.

Example: Grade 1 + 6 = age 7

Grade 9+6 = age 15

Grade 12+6=age 18

4. A learner must be admitted to grade 1 if he or she turns seven in the course of that calendar year. A learner who is younger than this age may not be admitted to grade 1.

5. A learner may be admitted to grade R only if he or she turns six in the course of that calendar year. Attendance of grade R is not compulsory."

In 2001, slight adjustments were made to include several clauses which allow for exceptions to be made, and a child to be admitted to school at an earlier age.

"4.(A) Despite paragraph 4, the Head of Department may allow a learner who wants to be admitted to grade 1, but who will not be turning 7 during the year of such admission, to be admitted at a lower age. This deviation by the Head of Department may only occur if, in the opinion of the Head of Department, reasonable grounds exist to show that such a learner is, based on educational principles, school ready for Grade 1 and it is in the best interest of such a learner to be admitted as an underage learner to a public school. The parent of the learner must show that the refusal to be admitted to a school will have a detrimental effect on the child's development."

However, just a year later the admissions' age range was updated, and the admission age of children into Grade 1 and Grade R shifted six months earlier. The Education Laws Amendment Act 50 of 2002 updates the 1998 Government Gazette notice to instead mandate and admissions age of:

2. Section 5 of the Schools Act is amended by the substitution for subsection (4) of the following subsection:

"(4) **(a)** The admission age of a **learner** to a **public** school to –

(i) **grade R** is age four turning five by 30 June in the year of admission;

(ii) **grade 1** is age five turning six by 30 June in the year of admission."

Additionally, the exceptions clause was altered slightly to the following:

(b) Subject to the availability of suitable school places and other educational resources, the **Head of Department** may admit a **learner** who–

(i) is under the age contemplated in paragraph **(a)** if good cause is shown; and (ii) complies with the criteria contemplated in paragraph (c).

(ii) criteria for the admission to a **public** school, at an age lower than the admission age, of an underage **learner** who complies with the criteria';

(ii) age requirements for different **grades** at a public school.

- (i) *it can be shown that exceptional circumstances exist which necessitate the admission of an underage **learner** because admission would be in his or her best interest; and*
- (ii) *the refusal to admit that **learner** would be severely detrimental to his or her development."*

This age range which mandated the age at which a child should be admitted into Grade R or Grade 1 has not been adjusted since 2002, and remains the official school admissions and policy in force today.

9.1.2 Compulsory school age

A different sub-section of the Act outlines the age range during which it is compulsory for a child to attend school. This first appeared in 1996, in Sub-section 3 and was stated as follows:

"(1) Subject to this Act and any applicable provincial law, every parent must cause every learner for whom he or she is responsible to attend a school from the first school day of the year in which such learner reaches the age of seven years until the last school day of the year in which such learner reaches the age of fifteen years or the ninth grade, whichever occurs first."

In the recent Basic Education Laws Amendment Act 32 of 2024, this was extended to incorporate Grade R. This reduced the compulsory school-entry age to 6 years so that parents are required to send their children to school at the latest:

"starting from grade R on the first school day of the year in which such learner reaches the age of six years"

9.2 School Age Policy Classification Process to identify Mid-year and Calendar Year schools

A four-step process was followed to classify each school in the dataset as either following a Mid-year age or a Calendar Year school-entry age policy.

Under the Mid-year policy, the majority of the cohort entering Grade 1 for the first time does so at the minimum possible age, namely, learners are between the ages of 5 ½ and 6 ½ years of age on the 1st of January in the year that they start Grade 1. Schools which follow the Calendar Year policy admit the majority of their Grade 1 first-time entrants in the latest possible 12-month period, namely learners are between the ages of 6 and 7 years old on the 1st of January in the year that they are first admitted into Grade 1.

Children born in the latter half of the year, from July to December, are required to start Grade 1 at six years of age. However, for children born between January and June, they have the option of starting at the age of 5 or 6. As a result, the default school-entry age for children born in the first half of the year determines whether a school is a Mid-year or Calendar Year school.

Step 1: Calculate the proportion of Grade 1 new entrants in each age category by school

The following age categories were created:

| Age range | Category |
|--------------------------------|--------------------|
| Less than 5.5 yrs | Underage |
| 5.5 yrs – 6 yrs | Mid-year only |
| 6yrs – 6.5 yrs | Both MY and CY |
| 6.5 – 7 yrs | Calendar Year only |
| Greater than or equal to 7 yrs | Overage |

For every school between 2017 and 2023, all Grade 1 first-time entrants were divided into these categories, and the proportion belonging to each category was calculated.

Step 2: Identify the most likely school-entry age policy by school and year (2017 – 2023)

Each school in each year was classified as being a Mid-year school if the proportion of learners following the Mid-year policy was higher than the proportion of learners entering during the Calendar Year policy age range. At least 60% of the learners adhered to the Mid-year policy. The same was done to identify Calendar Year schools.

Some schools remained unclassified, because they had large numbers of over- or underaged learners entering Grade 1. These remaining schools were then classified as Mid-year or Calendar Year by comparing the proportion of learners who entered in either Mid-year & Underaged to Calendar Year & Overaged categories.

Step 3: Assign the most likely school-entry age policy across all years to each school (2018 – 2023)

For each school, a count was conducted of the number of times the school was classified as “Mid-year” or “Calendar Year” between 2018 and 2023. The school was then classified as adhering to the policy which the school was allocated to in more than half of the years. If exactly 50% of the years that schools were designated as “Mid-year” and 50% of the time as “Calendar Year”, then the school was classified as being a “Mid-year” school.

The year 2017 was excluded because we were unable to identify which of the Grade 1's were repeaters, as this is the first year in the dataset. However, for schools that only appeared in the dataset in 2017, these were assigned the school-entry age policy identified in Step 2.

9.3 Data Drive District (DDD) Data completeness

The School Realities report is released annually by the Department of Basic Education and reports on the total number of schools, educators and learners at ordinary³⁴ schools in the country. In Table 20 the number of learners in the DDD dataset used is compared to those reported in the School Realities reports, this provides a view of how complete the data is by province and year. The first two years, 2017 and 2018 are the least complete. This is particularly acute in KwaZulu-Natal and Gauteng, where in 2017 just less than 70% of learners were captured in the DDD data in Gauteng and only 50-60% in KZN. This percentage rapidly rose and by 2019 at least 90% of learners were in the DDD dataset in each of the six provinces. The other 4 provinces, the Eastern Cape, Limpopo, Mpumalanga and the North West were already well above 90% completeness by 2017. A decrease in data completeness in 2020 across most provinces is likely due to COVID-19 pandemic disruptions.

From a representativeness perspective, it is worth noting that data is required from 2017 to create the 2018 cohort, which means that both the KZN and Gauteng samples are smaller and less representative of the provinces for the cohort that forms part of the analysis.

Table 20: Number of learners in the DDD dataset compared to those reported in School Realities

| Eastern Cape | | | | | | |
|--------------|------------------|------------------|------------------|------------------|-------------|------------|
| Year | School Realities | | DDD Data | | DDD % of SR | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 |
| 2017 | 166 883 | 472 437 | 165 614 | 470 161 | 99% | 100% |
| 2018 | 173 156 | 487 547 | 170 784 | 479 677 | 99% | 98% |
| 2019 | 162 880 | 483 022 | 162 117 | 479 845 | 100% | 99% |
| 2020 | 160 819 | 473 631 | 156 405 | 464 793 | 97% | 98% |
| 2021 | 154 750 | 459 090 | 154 061 | 457 297 | 100% | 100% |
| 2022 | 147 075 | 448 817 | 145 580 | 445 161 | 99% | 99% |
| 2023 | 136 776 | 435 227 | 134 303 | 429 156 | 98% | 99% |
| Total | 1 102 339 | 3 259 771 | 1 088 864 | 3 226 090 | 99% | 99% |

| Gauteng | | | | | | |
|--------------|------------------|------------------|------------------|------------------|-------------|------------|
| Year | School Realities | | DDD Data | | DDD % of SR | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 |
| 2017 | 239 952 | 662 943 | 159 093 | 459 492 | 66% | 69% |
| 2018 | 228 943 | 654 490 | 195 227 | 563 452 | 85% | 86% |
| 2019 | 228 622 | 656 964 | 205 441 | 592 711 | 90% | 90% |
| 2020 | 226 516 | 663 616 | 186 025 | 536 040 | 82% | 81% |
| 2021 | 225 162 | 666 163 | 197 735 | 588 074 | 88% | 88% |
| 2022 | 226 338 | 669 099 | 196 504 | 583 322 | 87% | 87% |
| 2023 | 222 041 | 661 978 | 183 155 | 547 157 | 82% | 83% |
| Total | 1 597 574 | 4 635 253 | 1 323 180 | 3 870 248 | 83% | 83% |

34 Excludes special needs schools.

| KwaZulu-Natal | | | | | | | |
|---------------|------------------|------------------|------------------|------------------|-------------|------------|--|
| Year | School Realities | | DDD Data | | DDD % of SR | | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | |
| 2017 | 254 644 | 718 234 | 141 877 | 402 072 | 56% | 56% | |
| 2018 | 243 894 | 702 502 | 180 124 | 511 129 | 74% | 73% | |
| 2019 | 239 933 | 698 291 | 216 851 | 626 970 | 90% | 90% | |
| 2020 | 237 531 | 687 417 | 194 301 | 565 208 | 82% | 82% | |
| 2021 | 231 204 | 680 363 | 193 259 | 569 471 | 84% | 84% | |
| 2022 | 217 020 | 674 528 | 192 879 | 594 482 | 89% | 88% | |
| 2023 | 208 583 | 662 273 | 185 216 | 586 526 | 89% | 89% | |
| Total | 1 632 809 | 4 823 608 | 1 304 507 | 3 855 858 | 80% | 80% | |

| Limpopo | | | | | | | |
|--------------|------------------|------------------|----------------|------------------|-------------|------------|--|
| Year | School Realities | | DDD Data | | DDD % of SR | | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | |
| 2017 | 150 632 | 437 856 | 139 973 | 409 628 | 93% | 94% | |
| 2018 | 148 603 | 430 122 | 146 359 | 424 709 | 98% | 99% | |
| 2019 | 147 177 | 437 608 | 145 576 | 431 950 | 99% | 99% | |
| 2020 | 146 910 | 441 581 | 142 330 | 428 412 | 97% | 97% | |
| 2021 | 142 757 | 442 206 | 141 074 | 436 330 | 99% | 99% | |
| 2022 | 135 001 | 435 008 | 133 493 | 429 976 | 99% | 99% | |
| 2023 | 133 725 | 424 119 | 131 762 | 417 951 | 99% | 99% | |
| Total | 1 004 805 | 3 048 500 | 980 567 | 2 978 956 | 98% | 98% | |

| Mpumalanga | | | | | | | |
|--------------|------------------|------------------|----------------|------------------|-------------|------------|--|
| Year | School Realities | | DDD Data | | DDD % of SR | | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | |
| 2017 | 96 726 | 282 983 | 90 152 | 270 213 | 93% | 95% | |
| 2018 | 94 075 | 267 219 | 95 232 | 270 660 | 101% | 101% | |
| 2019 | 95 507 | 275 399 | 93 875 | 271 281 | 98% | 99% | |
| 2020 | 94 427 | 275 161 | 90 737 | 264 326 | 96% | 96% | |
| 2021 | 92 049 | 278 365 | 90 687 | 270 970 | 99% | 97% | |
| 2022 | 90 525 | 276 315 | 87 955 | 269 808 | 97% | 98% | |
| 2023 | 87 515 | 273 044 | 84 468 | 263 917 | 97% | 97% | |
| Total | 650 824 | 1 928 486 | 633 106 | 1 881 175 | 97% | 98% | |

| North West | | | | | | | |
|--------------|------------------|------------------|----------------|------------------|-------------|------------|--|
| Year | School Realities | | DDD Data | | DDD % of SR | | |
| | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | Gr 1 | Gr 2-4 | |
| 2017 | 76 786 | 229 481 | 73 097 | 222 139 | 95% | 97% | |
| 2018 | 76 426 | 230 666 | 71 435 | 214 873 | 93% | 93% | |
| 2019 | 74 983 | 229 079 | 73 196 | 222 914 | 98% | 97% | |
| 2020 | 73 281 | 225 819 | 68 501 | 211 778 | 93% | 94% | |
| 2021 | 71 870 | 220 766 | 68 607 | 207 783 | 95% | 94% | |
| 2022 | 71 217 | 219 060 | 68 859 | 214 502 | 97% | 98% | |
| 2023 | 68 555 | 216 181 | 62 678 | 195 952 | 91% | 91% | |
| Total | 513 118 | 1 571 052 | 486 373 | 1 489 941 | 95% | 95% | |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW and School Realities Reports 2017 – 2023

Table 21: Table comparing sample size of the Grade 1 2018 sample in the DDD cleaned and ordered relative to the original DDD data

| Year | EC | GP | KZN | LP | MP | NW | All 6 Prov. |
|--------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|
| 2017 | 100.0% | 99.7% | 100.2% | 100.1% | 100.1% | 99.8% | 100.0% |
| 2018 | 100.3% | 100.0% | 100.3% | 100.1% | 100.1% | 99.9% | 100.1% |
| 2019 | 100.5% | 100.0% | 100.1% | 100.1% | 100.0% | 99.8% | 100.1% |
| 2020 | 100.9% | 100.0% | 100.2% | 100.1% | 100.2% | 99.9% | 100.2% |
| 2021 | 100.7% | 100.0% | 100.2% | 100.0% | 100.3% | 99.9% | 100.2% |
| 2022 | 101.0% | 100.0% | 100.2% | 100.0% | 100.2% | 99.8% | 100.2% |
| 2023 | 100.8% | 100.0% | 100.1% | 100.0% | 100.1% | 99.8% | 100.2% |
| Total | 100.6% | 99.9% | 100.2% | 100.1% | 100.1% | 99.8% | 100.2% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GP, KZN, LP, MP and NW and School Realities Reports 2017 – 2023. Note: Numbers above 100%, can occur if there were cases where a learner identifier appears to have been assigned to two different individuals (eg. in Grade 2 and 4 at the same time). They were then assigned two separate new identifiers.

Table 22: Number of observations for learners in Grade 1 in 2018 across samples

| | School Realities | DDD Clean. & Ord. | School Panel | Gr 1&4 Learner panel (w. rep.) |
|---------------------------------------|------------------|-------------------|----------------|--------------------------------|
| # of observation | | | | |
| EC | 173 156 | 171 282 | 157 191 | 138 762 |
| GT | 228 943 | 195 164 | 141 934 | 122 203 |
| KZN | 243 894 | 180 717 | 114 434 | 100 543 |
| LP | 148 603 | 146 499 | 139 506 | 130 320 |
| MP | 94 075 | 95 305 | 89 054 | 81 739 |
| NW | 76 426 | 71 333 | 66 284 | 59 915 |
| Total | 965 097 | 860 300 | 708 403 | 633 482 |
| % relative to School Realities | | | | |
| EC | 100.0% | 98.9% | 90.8% | 80.1% |
| GT | 100.0% | 85.2% | 62.0% | 53.4% |
| KZN | 100.0% | 74.1% | 46.9% | 41.2% |
| LP | 100.0% | 98.6% | 93.9% | 87.7% |
| MP | 100.0% | 101.3% | 94.7% | 86.9% |
| NW | 100.0% | 93.3% | 86.7% | 78.4% |
| Total | 100.0% | 89.1% | 73.4% | 65.6% |
| % relative to DDD cleaned | | | | |
| EC | 101.1% | 100.0% | 91.8% | 81.0% |
| GT | 117.3% | 100.0% | 72.7% | 62.6% |
| KZN | 135.0% | 100.0% | 63.3% | 55.6% |
| LP | 101.4% | 100.0% | 95.2% | 89.0% |
| MP | 98.7% | 100.0% | 93.4% | 85.8% |
| NW | 107.1% | 100.0% | 92.9% | 84.0% |
| Total | 112.2% | 100.0% | 82.3% | 73.6% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GP, KZN, LP, MP and NW and School Realities Reports 2017 – 2023. Note: Last column includes Grade 1 repeaters to allow for comparability with the cleaned DDD and School Panel datasets.

Table 23: Age and gender for learners in the various Grade 1 2018 sample

| | DDD Clean. & Ord. | School Panel | Gr 1&4 Learner panel (w. rep.) |
|------------------------|-------------------|--------------|--------------------------------|
| Mean age (Gr 1) | | | |
| EC | 6.34 | 6.34 | 6.33 |
| GT | 6.29 | 6.27 | 6.26 |
| KZN | 6.18 | 6.18 | 6.17 |
| LP | 6.13 | 6.13 | 6.12 |
| MP | 6.26 | 6.26 | 6.25 |
| NW | 6.32 | 6.32 | 6.30 |
| Total | 6.25 | 6.25 | 6.23 |
| % female (Gr 1) | | | |
| EC | 47.0% | 47.0% | 47.3% |
| GT | 47.9% | 47.7% | 48.2% |
| KZN | 47.4% | 47.6% | 47.7% |
| LP | 47.7% | 47.7% | 47.8% |
| MP | 48.0% | 48.0% | 48.3% |
| NW | 47.9% | 48.0% | 48.5% |
| Total | 47.6% | 47.6% | 47.9% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Note: Last column includes Grade 1 repeaters to allow for comparability with the cleaned DDD and School Panel datasets.

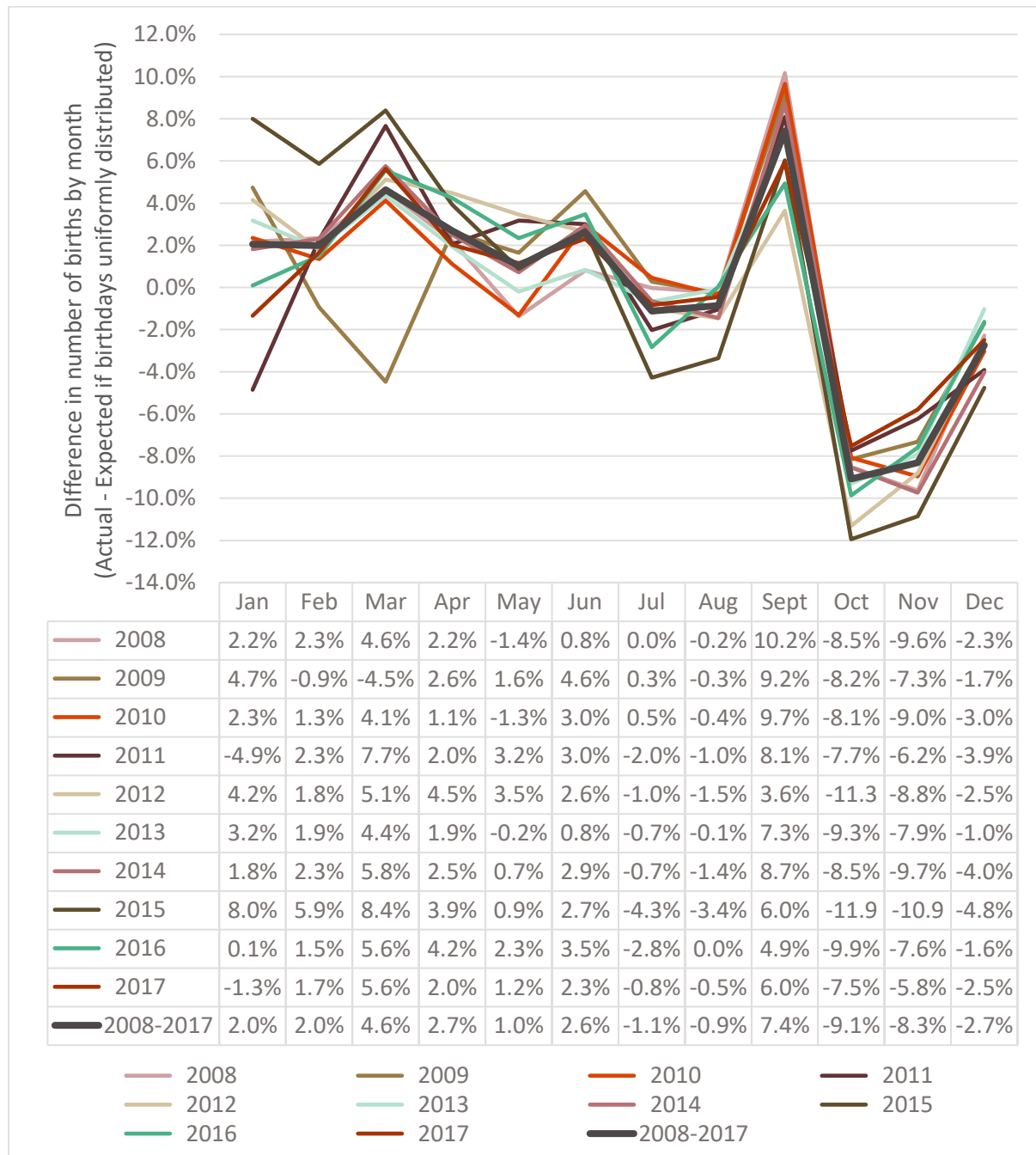
Table 24: Percentage of observations remaining in the school and learner panels relative to the DDD cleaned and ordered dataset, by learner characteristics

| | % of observations relative to DDD Cl & Ord. data | |
|--------------------------|--|--------------------------------|
| | School Panel | Gr 1&4 Learner panel (w. rep.) |
| Race | | |
| African/Black | 83% | 74% |
| Asian/Indian | 67% | 59% |
| Coloured | 82% | 76% |
| Other | 71% | 55% |
| White | 70% | 55% |
| Home Language | | |
| Afrikaans | 81% | 69% |
| English | 66% | 56% |
| IsiNdebele | 82% | 76% |
| IsiXhosa | 89% | 78% |
| IsiZulu | 72% | 63% |
| SePedi | 88% | 82% |
| SeSotho | 87% | 76% |
| SeTswana | 87% | 79% |
| SiSwati | 91% | 84% |
| Sign Language | 67% | 49% |
| TshiVenda | 94% | 88% |
| XiTsonga | 92% | 84% |
| School Sector | | |
| Public | 83% | 75% |
| Indep. | 65% | 51% |
| School Age Policy | | |
| Mid-year | 83% | 75% |
| Calendar Year | 76% | 66% |
| Total | 82% | 74% |

Source: DDD learner panel, 2017 – 2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Notes: One of the categories for Home Language allows for a combination to be entered as "Afr/Eng/Xhosa". What this combination refers to is not clear, and as it makes up less than 1% of the sample, it is not reported here.

9.4 Seasonal Patterns of Births in South Africa

Figure 13: Number of actual births by month relative to a uniform distribution of births in South Africa, 2008 - 2017



Source: Statistics South Africa Birth Registration Data, own calculations.

9.5 Regression results by province

Table 25: Relative age effect on Grade 1 outcomes, by province for the Grade 1 2018 cohort

| Mid-year | | | | | | | |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | EC | GP | KZN | LP | MP | NW | All |
| Gr 1 Rep | | | | | | | |
| Coeff | -0.217*** | -0.186*** | -0.164*** | -0.126*** | -0.162*** | -0.128*** | -0.160*** |
| SE | (0.010) | (0.005) | (0.005) | (0.003) | (0.006) | (0.007) | (0.002) |
| Mean | 0.183 | 0.128 | 0.137 | 0.094 | 0.135 | 0.105 | 0.131 |
| N | 96 169 | 90 712 | 84 191 | 115 911 | 61 128 | 40 336 | 488 447 |
| HL mark | | | | | | | |
| Coeff | 12.149*** | 13.130*** | 13.239*** | 12.199*** | 13.303*** | 11.222*** | 12.638*** |
| SE | (0.380) | (0.236) | (0.252) | (0.161) | (0.294) | (0.355) | (0.104) |
| Mean | 59.829 | 64.866 | 65.926 | 67.966 | 66.744 | 67.280 | 65.225 |
| N | 96 121 | 90 538 | 83 936 | 115 823 | 60 822 | 40 158 | 487 398 |
| MTH mark | | | | | | | |
| Coeff | 13.175*** | 13.327*** | 13.349*** | 13.054*** | 13.290*** | 11.717*** | 13.100*** |
| SE | (0.402) | (0.241) | (0.246) | (0.175) | (0.301) | (0.380) | (0.107) |
| Mean | 62.671 | 69.944 | 70.271 | 71.517 | 73.199 | 71.738 | 69.496 |
| N | 96 154 | 90 696 | 84 145 | 115 898 | 61 123 | 40 157 | 488 173 |
| FAL mark | | | | | | | |
| Coeff | 11.388*** | 11.545*** | 11.446*** | 11.242*** | 10.943*** | 10.003*** | 11.215*** |
| SE | (0.371) | (0.219) | (0.226) | (0.155) | (0.260) | (0.330) | (0.096) |
| Mean | 59.409 | 66.287 | 66.044 | 68.587 | 69.725 | 69.744 | 66.153 |
| N | 95 502 | 89 914 | 83 786 | 115 446 | 60 392 | 40 240 | 485 280 |
| Calendar Year | | | | | | | |
| | EC | GP | KZN | LP | MP | NW | All |
| Gr 1 Rep | | | | | | | |
| Coeff | -0.130*** | -0.092*** | -0.058*** | -0.059*** | -0.102*** | -0.098*** | -0.102*** |
| SE | (0.015) | (0.010) | (0.020) | (0.021) | (0.013) | (0.011) | (0.006) |
| Mean | 0.141 | 0.074 | 0.047 | 0.058 | 0.094 | 0.089 | 0.097 |
| N | 19 090 | 18 629 | 2 730 | 2 763 | 10 979 | 13 877 | 68 068 |
| HL mark | | | | | | | |
| Coeff | 8.872*** | 9.183*** | 8.744*** | 8.471*** | 10.390*** | 9.890*** | 9.461*** |
| SE | (0.621) | (0.491) | (1.209) | (1.261) | (0.679) | (0.565) | (0.276) |
| Mean | 62.250 | 69.754 | 71.003 | 68.771 | 67.942 | 68.883 | 67.195 |
| N | 18 976 | 18 607 | 2 710 | 2 763 | 10 979 | 13 834 | 67 869 |
| MTH mark | | | | | | | |
| Coeff | 9.671*** | 9.781*** | 8.948*** | 9.177*** | 11.645*** | 10.057*** | 10.114*** |
| SE | (0.660) | (0.501) | (1.256) | (1.290) | (0.684) | (0.582) | (0.286) |
| Mean | 66.988 | 74.282 | 76.053 | 72.970 | 73.967 | 73.739 | 72.091 |
| N | 19 087 | 18 628 | 2 710 | 2 763 | 10 979 | 13 835 | 68 002 |
| FAL mark | | | | | | | |
| Coeff | 7.364*** | 7.601*** | 8.201*** | 8.004*** | 8.565*** | 8.755*** | 8.048*** |
| SE | (0.598) | (0.442) | (1.188) | (1.194) | (0.604) | (0.514) | (0.254) |
| Mean | 63.542 | 72.054 | 73.356 | 69.950 | 70.236 | 70.918 | 69.146 |
| N | 18 521 | 18 592 | 2 705 | 2 763 | 10 557 | 13 876 | 67 014 |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GP, KZN, LP, MP and NW.

Table 26: Relative age effect on Grade 4 outcomes, by province for the Grade 1 2018 cohort

| | Mid-year | | | | | | |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | EC | GP | KZN | LP | MP | NW | All |
| FP Rep | | | | | | | |
| Coeff | -0.263*** | -0.243*** | -0.211*** | -0.190*** | -0.216*** | -0.184*** | -0.215*** |
| SE | (0.012) | (0.006) | (0.006) | (0.004) | (0.007) | (0.011) | (0.003) |
| Mean | 0.333 | 0.249 | 0.243 | 0.205 | 0.247 | 0.259 | 0.254 |
| N | 95 558 | 91 862 | 84 120 | 114 790 | 61 472 | 40 486 | 488 288 |
| Gr4 Rep | | | | | | | |
| Coeff | -0.057*** | -0.027*** | -0.033*** | -0.050*** | -0.038*** | -0.049*** | -0.042*** |
| SE | (0.007) | (0.003) | (0.003) | (0.003) | (0.004) | (0.007) | (0.002) |
| Mean | 0.086 | 0.049 | 0.054 | 0.082 | 0.056 | 0.110 | 0.071 |
| N | 95 558 | 91 862 | 84 120 | 114 790 | 61 472 | 40 486 | 488 288 |
| HL mark | | | | | | | |
| Coeff | 3.731*** | 4.853*** | 4.775*** | 6.072*** | 4.962*** | 5.059*** | 5.090*** |
| SE | (0.332) | (0.205) | (0.201) | (0.161) | (0.254) | (0.365) | (0.093) |
| Mean | 62.257 | 66.445 | 66.478 | 66.587 | 67.142 | 64.023 | 65.554 |
| N | 95 187 | 91 801 | 83 955 | 114 777 | 61 430 | 40 405 | 487 555 |
| MTH mark | | | | | | | |
| Coeff | 4.860*** | 5.630*** | 5.960*** | 6.638*** | 5.403*** | 5.964*** | 5.886*** |
| SE | (0.370) | (0.213) | (0.217) | (0.163) | (0.250) | (0.367) | (0.097) |
| Mean | 59.047 | 63.693 | 63.312 | 63.307 | 64.397 | 61.806 | 62.562 |
| N | 95 180 | 91 858 | 84 054 | 114 751 | 61 412 | 40 481 | 487 736 |
| FAL mark | | | | | | | |
| Coeff | 3.914*** | 5.221*** | 5.350*** | 6.637*** | 5.291*** | 5.450*** | 5.531*** |
| SE | (0.345) | (0.211) | (0.213) | (0.166) | (0.255) | (0.364) | (0.096) |
| Mean | 57.935 | 62.406 | 61.513 | 61.636 | 62.569 | 59.296 | 60.962 |
| N | 95 003 | 91 853 | 84 040 | 114 759 | 61 386 | 40 454 | 487 495 |

| | Calendar Year | | | | | | |
|-----------------|---------------|-----------|----------|----------|-----------|-----------|-----------|
| | EC | GP | KZN | LP | MP | NW | All |
| FP Rep | | | | | | | |
| Coeff | -0.144*** | -0.116*** | -0.052* | -0.061** | -0.103*** | -0.108*** | -0.114*** |
| SE | (0.018) | (0.012) | (0.028) | (0.028) | (0.016) | (0.015) | (0.007) |
| Mean | 0.243 | 0.123 | 0.098 | 0.108 | 0.164 | 0.204 | 0.178 |
| N | 18 793 | 18 502 | 2 715 | 2 702 | 10 696 | 13 547 | 66 955 |
| Gr4 Rep | | | | | | | |
| Coeff | -0.016 | -0.010** | -0.025* | -0.006 | -0.008 | -0.029*** | -0.016*** |
| SE | (0.010) | (0.005) | (0.014) | (0.017) | (0.009) | (0.011) | (0.004) |
| Mean | 0.057 | 0.016 | 0.028 | 0.034 | 0.042 | 0.083 | 0.046 |
| N | 18 793 | 18 502 | 2 715 | 2 702 | 10 696 | 13 547 | 66 955 |
| HL mark | | | | | | | |
| Coeff | 4.713*** | 4.335*** | 5.046*** | 4.124*** | 5.396*** | 3.986*** | 4.561*** |
| SE | (0.597) | (0.408) | (1.110) | (1.116) | (0.628) | (0.560) | (0.256) |
| Mean | 64.535 | 69.313 | 70.148 | 69.214 | 67.064 | 64.126 | 66.594 |
| N | 18 766 | 18 502 | 2 715 | 2 702 | 10 696 | 13 544 | 66 925 |
| MTH mark | | | | | | | |
| Coeff | 4.596*** | 4.942*** | 4.193*** | 5.150*** | 6.166*** | 4.607*** | 4.992*** |
| SE | (0.689) | (0.525) | (1.420) | (1.380) | (0.718) | (0.606) | (0.300) |
| Mean | 62.389 | 67.298 | 67.786 | 66.783 | 64.436 | 61.757 | 64.342 |
| N | 18 753 | 18 501 | 2 715 | 2 702 | 10 617 | 13 547 | 66 835 |
| FAL mark | | | | | | | |
| Coeff | 4.550*** | 5.054*** | 4.084*** | 4.311*** | 5.626*** | 4.837*** | 4.930*** |
| SE | (0.612) | (0.447) | (1.236) | (1.203) | (0.646) | (0.578) | (0.268) |
| Mean | 61.729 | 66.713 | 69.024 | 66.793 | 63.559 | 60.953 | 63.746 |
| N | 18 645 | 18 501 | 2 715 | 2 702 | 10 667 | 13 544 | 66 774 |

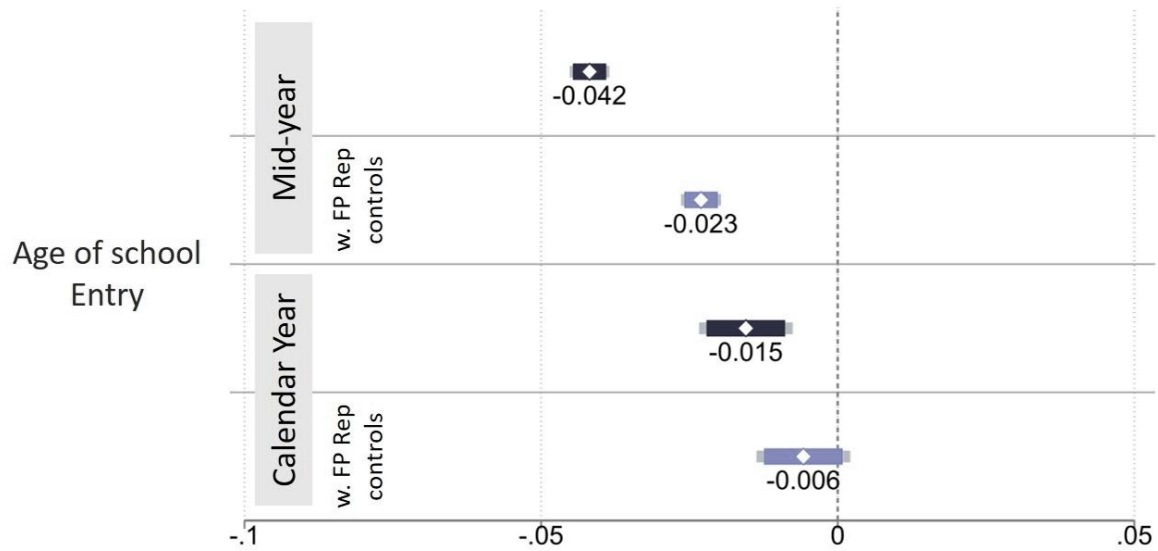
Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW.

9.6 Indirect effect of school-entry age on outcomes via repetition

The influence of an additional year of schooling and an increase in age by a calendar year cannot simply be accounted for by adding controls for Foundation Phase repetition, as illustrated in Figure 14. Doing so would introduce new bias: repeaters would be compared with repeaters, and non-repeaters with non-repeaters, even though younger and older entrants were not randomly sorted into these groups rather, ability and school-entry age influenced which learners repeated³⁵. Unobserved learner characteristics such as academic ability, motivation, health or family situation also act as confounders for repetition (as well as for school-entry age).

35 For learners who were initially among the youngest in the cohort the rate of repetition was much higher, suggesting that their underlying academic ability might also be higher relative to older repeaters. Similarly, for the non-repeaters, fewer of the younger learners (only the most academically ready) were promoted. Thus, this group is also likely to be an academically stronger group on average relative to the older learners. Therefore, by including repetition this actually makes the groups being compared less comparable.

Figure 14: Relative age effect on Grade 4 outcomes with and without controls for FP Repetition, for the Grade 1 2018 cohort



Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Sample size is 488 492 for the Mid-year policy and 68 070 for the Calendar Year policy. Used the command `ivreghdfe` in Stata to run an IV regression with expected school-entry age as an instrument for actual age of first-time Grade 1 entry with school fixed effects and controlling for gender and Grade R attendance.

9.7 Upward and Downward Rank Probabilities ($\delta = 0$ & $\delta = 0.05$)

The tables below show the upward and downward rank mobilities at two different deltas (δ). The first, where $\delta = 0$, looks at the cases where any upward or downward movement is counted, whilst the case of $\delta = 0.05$ only counts learners that move at least 5 percentiles up or down in ranking between Grade 1 and Grade 4.

Table 27: Upward Rank Probabilities ($\delta= 0$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| Subject | Gr 1 Subject Quartile | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|----------|----------------|-----|-----|----------|
| | | Birth Quartile | | | | Birth Quartile | | | |
| | | Oldest | | | Youngest | Oldest | | | Youngest |
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| HL | 1 | 69% | 70% | 72% | 74% | 69% | 69% | 70% | 72% |
| HL | 2 | 49% | 51% | 53% | 55% | 50% | 51% | 51% | 56% |
| HL | 3 | 40% | 42% | 42% | 43% | 42% | 43% | 44% | 45% |
| HL | 4 | 29% | 30% | 30% | 31% | 31% | 32% | 34% | 34% |
| MTH | 1 | 70% | 71% | 73% | 75% | 68% | 69% | 70% | 73% |
| MTH | 2 | 51% | 52% | 54% | 55% | 51% | 54% | 54% | 57% |
| MTH | 3 | 41% | 41% | 43% | 42% | 42% | 44% | 45% | 47% |
| MTH | 4 | 29% | 29% | 30% | 30% | 30% | 32% | 31% | 34% |
| FAL | 1 | 71% | 72% | 74% | 75% | 73% | 74% | 73% | 76% |
| FAL | 2 | 53% | 53% | 55% | 56% | 57% | 56% | 56% | 57% |
| FAL | 3 | 42% | 42% | 43% | 43% | 44% | 44% | 46% | 46% |
| FAL | 4 | 29% | 29% | 29% | 29% | 29% | 29% | 29% | 30% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in [0; 1 - \delta)$ are included in the analysis.

Table 28: Downward Rank Probabilities ($\delta= 0$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| Subject | Gr 1 Subject Quartile | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|----------|----------------|-----|-----|----------|
| | | Birth Quartile | | | | Birth Quartile | | | |
| | | Oldest | | | Youngest | Oldest | | | Youngest |
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| HL | 1 | 30% | 29% | 28% | 25% | 31% | 31% | 30% | 28% |
| HL | 2 | 50% | 49% | 46% | 45% | 50% | 49% | 49% | 44% |
| HL | 3 | 60% | 58% | 57% | 57% | 58% | 57% | 56% | 55% |
| HL | 4 | 71% | 70% | 70% | 69% | 70% | 68% | 67% | 66% |
| MTH | 1 | 30% | 29% | 27% | 25% | 31% | 31% | 29% | 27% |
| MTH | 2 | 49% | 47% | 46% | 45% | 48% | 46% | 45% | 42% |
| MTH | 3 | 59% | 59% | 57% | 58% | 58% | 56% | 55% | 53% |
| MTH | 4 | 71% | 71% | 70% | 70% | 70% | 68% | 69% | 66% |
| FAL | 1 | 29% | 27% | 26% | 24% | 27% | 26% | 27% | 24% |
| FAL | 2 | 47% | 47% | 44% | 44% | 43% | 43% | 43% | 43% |
| FAL | 3 | 58% | 58% | 57% | 57% | 56% | 55% | 54% | 54% |
| FAL | 4 | 72% | 71% | 71% | 71% | 71% | 70% | 71% | 70% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in (0 + \delta; 1]$ are included in the analysis.

Table 29: Upward Rank Probabilities ($\delta= 0.05$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| | | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|----------|----------------|-----|-----|----------|
| Subject | Gr 1 Subject Quartile | Birth Quartile | | | | Birth Quartile | | | |
| | | Oldest | | | Youngest | Oldest | | | Youngest |
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| HL | 1 | 58% | 59% | 61% | 63% | 57% | 57% | 57% | 60% |
| HL | 2 | 43% | 45% | 47% | 48% | 43% | 44% | 44% | 48% |
| HL | 3 | 33% | 35% | 35% | 36% | 35% | 36% | 37% | 38% |
| HL | 4 | 19% | 20% | 21% | 21% | 20% | 21% | 22% | 23% |
| MTH | 1 | 59% | 59% | 61% | 63% | 57% | 57% | 57% | 61% |
| MTH | 2 | 44% | 46% | 47% | 48% | 44% | 46% | 47% | 49% |
| MTH | 3 | 34% | 34% | 36% | 36% | 35% | 36% | 37% | 40% |
| MTH | 4 | 20% | 20% | 21% | 21% | 20% | 22% | 22% | 24% |
| FAL | 1 | 60% | 61% | 63% | 64% | 62% | 63% | 63% | 64% |
| FAL | 2 | 46% | 47% | 49% | 49% | 49% | 49% | 50% | 50% |
| FAL | 3 | 35% | 36% | 36% | 36% | 38% | 38% | 39% | 39% |
| FAL | 4 | 20% | 20% | 21% | 20% | 21% | 20% | 21% | 21% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in [0; 1 - \delta)$ are included in the analysis.

Table 30: Downward Rank Probabilities ($\delta= 0.05$) between Grade 1 and 4 for HL, MTH and FAL subject mark quartiles and birth quartiles

| | | Mid-year | | | | Calendar Year | | | |
|---------|-----------------------|----------------|-----|-----|----------|----------------|-----|-----|----------|
| Subject | Gr 1 Subject Quartile | Birth Quartile | | | | Birth Quartile | | | |
| | | Oldest | | | Youngest | Oldest | | | Youngest |
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| HL | 1 | 21% | 20% | 19% | 18% | 22% | 23% | 21% | 18% |
| HL | 2 | 44% | 42% | 40% | 38% | 42% | 42% | 40% | 37% |
| HL | 3 | 54% | 51% | 50% | 50% | 51% | 49% | 48% | 48% |
| HL | 4 | 58% | 58% | 57% | 57% | 55% | 54% | 53% | 53% |
| MTH | 1 | 21% | 20% | 19% | 17% | 22% | 21% | 20% | 18% |
| MTH | 2 | 42% | 41% | 39% | 38% | 41% | 39% | 39% | 36% |
| MTH | 3 | 52% | 52% | 50% | 51% | 51% | 49% | 48% | 45% |
| MTH | 4 | 58% | 58% | 58% | 59% | 57% | 55% | 57% | 53% |
| FAL | 1 | 20% | 19% | 18% | 17% | 19% | 18% | 18% | 17% |
| FAL | 2 | 40% | 40% | 38% | 38% | 37% | 37% | 36% | 37% |
| FAL | 3 | 52% | 51% | 51% | 50% | 49% | 49% | 47% | 47% |
| FAL | 4 | 60% | 59% | 59% | 60% | 59% | 59% | 59% | 59% |

Source: The 2018 Grade 1 first-time entrants' cohort balanced panel for Grades 1 and 4, from the DDD dataset 2017–2023 for 6 provinces: EC, GT, KZN, LP, MP and NW. Percentiles were calculated at the school level for all learners within that grade, including those from outside of the 2018 Grade 1 cohort. Only learners with percentiles in the range $\in (0 + \delta; 1]$ are included in the analysis.