Prediction and Prevention:

Reducing Repetition & Dropout of Primary Students in Rwanda

Ben Brockman & Vincent Vanderputten

SYPA Section Leader: Professor Rema Hanna

SYPA Adviser: Professor Asim Khwaja

Submitted March 11th, 2017

In partial fulfillment of the requirements for the degree of

Master in Public Administration in International Development,

John F. Kennedy School of Government, Harvard University

Acknowledgements:

First, to the team in Rwanda - we would like to thank Dimitri Stoelinga at Laterite for the initial inspiration for this project, gracious hosting in Kigali, and support throughout the entire SYPA process. We would also like to thank Patrick Hitayezu at Laterite for his support during our visit to Kigali in January and help in arranging meetings with Ministry of Education officials. We also thank the officials at the Ministry of Education and Rwandan Education Board who took the time to meet us and share their insights. We hope the results of project will contribute to your broader efforts in reducing repetition and dropout across the education system in Rwanda.

Second, we would like to thank our advisors Professor Rema Hanna and Professor Asim Khwaja for supporting us throughout the process. We appreciated the opportunity to pursue a, in some ways, "non-traditional" SYPA, focused on the data science methods we have been exploring this year. Finally, we would like to thank the HKS Degree Programs Student Affairs (DPSA) for the travel support that made our trip to Kigali in January 2017 possible.

Table of Content

1 Executive Summary

Rwanda has made remarkable progress in recent years in achieving near universal enrollment in primary education. However, Rwandan primary students continue to repeat grades and drop out of schools at high rates. From 2011 to 2015, the average dropout rate per year for a primary school student was 10% and the average repetition rate per year was 16.5%.

This policy brief explores whether "predictive analytics" based on short surveys with students and the parents can be used to identify students at high risk of repetition or dropout. This approach has two potential benefits: 1) early identification allows preventative measures to take place and 2) accurate prediction allows resources to be better targeted.

Using EICV data we create an 11 question "risk scorecard" for repetition that can be administered by school officials at Umuganda community meetings. The scorecard includes questions on basic literacy and numeracy levels, overaging for grade, and both pupil and household characteristics. With our scorecard, we are able to correctly identify close to 60% of truly at risk students, while only misclassifying 30% of truly not at-risk students.

Accurately identifying which students are at highest risk of repetition or dropout does not by itself solve the problem. We explore two ways scorecards could be incorporated into policy:

1) **Scorecards + Imihigo Contracts:** In this approach scorecards would be used to provide teachers and school officials information on which are the at-risk students to focus their mitigation efforts on. Imihigo targets around repetition or dropout can incentivize action.
2) **Scorecards + Targeted Programs:** In this approach scorecards would be used to select which students are eligible for a specific remedial program focused on preventing dropout and / or repetition. This program would be designed by MINEDUC and its partners to focus on root causes of the dropout issue.

We remain agnostic as to which proposal (or a hybrid of the two) has higher potential. We propose that the MINEDUC test these options using an experimental pilot before scaling up any specific approach. Due to data availability constraints, our research currently focuses on predicting repetition. However, following the completion of the MINEDUC/Laterite dropout and repetition study in May 2017, it will be possible to replicate this study for predicting dropout.

2 Research Background and Report Overview

This analysis has been prepared for senior policy makers at the Rwandan Ministry of Education and proposes new approaches to reducing repetition and dropout among primary school student. This work is a complementary report to a more extensive multi-year collaboration between Laterite and the Ministry of Education (MINEDUC), focused on the same set of policy topics.

This analysis differs in approach from Laterites' primary analysis which focuses on traditional data collection and regression analysis of potential causes of dropout and repetition. This report conducts an exploratory machine learning exercise to assess the viability of using predictive analytics to better target interventions to reduce dropout and repetition. Due to the timeframe of this policy analysis, the data Laterite will be collecting is not yet available, limiting our analysis to national survey data from the most recent EICV.

Our analysis is further confined to only the risk of student repetition, not dropouts, due to limits in the EICV data. We believe our analysis methods and recommendations on how to use predictive analysis in addressing repetition will be transferable to analyzing dropout when the relevant data becomes available. In the interim, we believe analyzing primary student repetition is an important policy goal in and of itself, demonstrated by our review of the literature and the budget pressures it places on the government.

The paper broadly proceeds as follows. First, we define key terms, examine the current state of dropout and repetition in Rwanda, review the literature, and describe unique features of the Rwandan education policy environment most relevant to our analysis. Second, we summarize our analytic approach in general terms and identify the benefits and drawbacks to our predictive approach. Third, we describe the methodology we followed to create a "predictive scorecard": a basic tool to be used by teachers or community leaders to identify which students are most at risk of repeating school. Fourth, we discuss how this scorecard can fit into broader policy frameworks and analyze implementation options. Finally, for the more technically inclined reader, we include an appendix highlighting the details of our machine learning approach.

3 Repetition and Dropout: Problem Definition and Background

Each student ends each school year with one of three outcomes - 'Promotion', 'Repetition', or 'Dropout'. 'Promotion' is defined as any case where a pupil acquires all the skills and knowledge required to pass exams, gets promoted to the next grade, and effectively enrolls in the next grade. The other two outcomes, repetition and dropout, are our outcomes of interest and are defined below.

Figure 1: Key Definitions

Defining Repetition: any case where a student enrolls in the same grade as the previous year. This can be: ● involuntary, when the school decides to hold back a pupil due to low learning outcomes ● a voluntary household-level decision, for logistic or financial reasons, for example when a new school is too far or too expensive, but they still want the child to be enrolled.
Defining Dropout: any case where a student was in school last year, but does not enroll the subsequent year. Some possible reasons are: ● A pupil fails exams and the pupil or household decides to not repeat but dropout instead. ● Household factors (financial, family obligations, etc.) make staying in school undesirable.

3.1 The Downsides of Repetition and Dropout

Students repeating grades or dropping from school come at a cost. This cost to society either comes as a direct cost (e.g. the financial cost of paying for additional years of schooling when pupils repeat) or as an opportunity cost (e.g. unrealized economic returns at the individual level[1]). Moreover, repeating grades is not necessarily in the interest of the child and his/her learning. The downsides to dropping out of primary school are self-evident and should be prevented. This section further explores the downsides to repetition which are less transparent.

[1] In fact, the World Bank estimated Rwanda to have the highest returns to (primary) education in the world, at 33%. See: Montenegro and Patrinos (2014)

At the Micro level

Whether repetition is overall desirable or whether it has a negative impact on an individual's academic and nonacademic outcomes is a topic of debate, and the effect depends on the context and system[2]. In theory, being held back can help a student catch up to classmates before advancing grades. This would allow the student to solidify understanding and build a better foundation for future learning. Some researchers indeed find benefits to short term learning outcome[3]. Others however find that early-grade repeating does not improve learning outcomes relative to being promoted[4] and several studies find medium term[5] and long term[6] negative impacts, even in pre-primary ages[7]. In summary, while short term learning improves, as students see the same material a second time, most evidence seems to indicate that repetition is not only ineffective at improving longer term learning outcomes, but even counterproductive.

In the current context, Rwanda has high enrollment rates in primary, in part because of a policy preference of repeating over dropping out[8]. However, learning outcomes have not kept pace with increases in the enrollment rate[9]. This is line with findings from the literature that repeating grades does not necessarily improve learning outcomes.

Beyond potential learning impacts of repetition, there are other undesirable effects. While repetition and over-aging do not necessarily have direct negative peer effects[10], repetition and dropout are interlinked: grade repetition have been shown to lead to increased dropout rates[11], especially for girls.[12] For two students with similar ability, background and context, the one

---

[2] Ikeda and García (2013)
[3] Schwerdt and West (2012)
[4] Glick and Sahn (2010)
[5] Manacorda (2010)
[6] Brophy (2006)
[7] Morgan (2006)
[8] Local government and leaders actively engage in contacting the households with students who dropped out in primary to get them enrolled again in school. This preference is also reflected in a recent decrease in dropout rate at the cost of an increased repetition rate, depicted in figure A1 in Appendix 10.1.
[9] While the enrollment rate is high, literacy rates are still low. While part of this low literacy rate is due to a lag in schooling effect on the average literacy rate. However, pupils' learning outcomes are still low for the current generation of primary pupils, despite a high-level curriculum. See for example the Learning Assessments in Rwandan Schools (LARS) outcomes (see: *Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region: Rwanda Country Case Study*, 2016)
[10] Jones (2016)
[11] Chihana et al. (2017)
[12] Grant and Hallman (2008), Nekatibeb (2002)

repeating will more likely eventually dropout than the one who was promoted[13]. Hence, we expect reduced repetition will indirectly improve dropout rates.

At the Macro-level

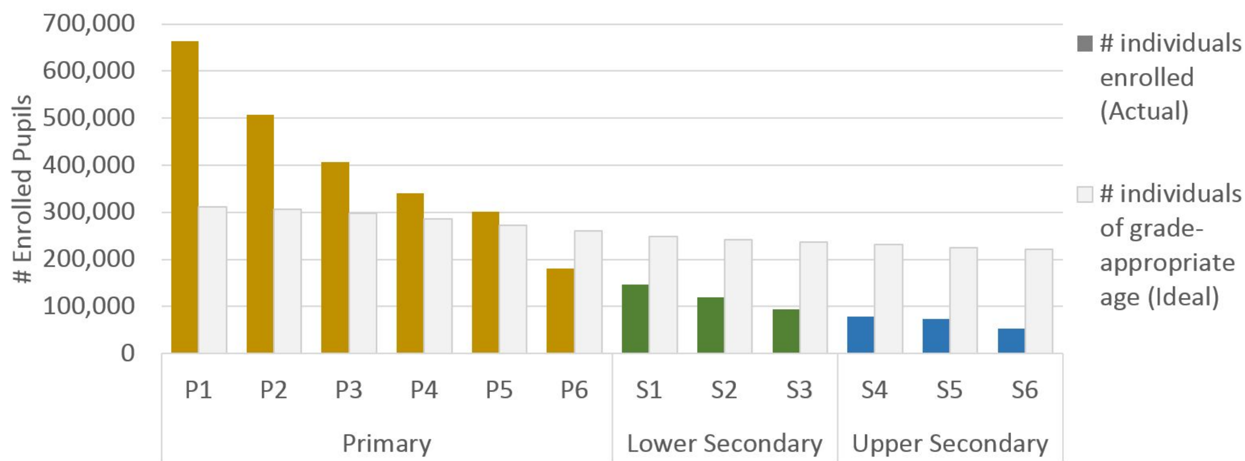
On a macro-level, repetition imposes large financial costs on the education system and can be considered an inefficiency in the schooling process. For example, a schooling system where pupils finish P6 in 8 years (i.e. repeating twice) requires a budget roughly 33% higher than a system where pupils remain on-track and finish P6 on times. These potential cost savings could then go into improving the quality of education in other ways. If the budget does not increase, inputs-per-students (e.g. student-teacher ratio) will be decrease. This will likely worsen learning outcomes and further increases repetition.

3.2 Dropout and Repetition in the Rwandan Education Sector

Rwanda has made much progress in delivering primary education: the net enrollment rate (NER) in primary increased from 62.5% in 1990 to 91.7% in 2010 and to 96.9% in 2015. Thus, almost all 7- to 12-year-olds today are enrolled in (primary) schooling. However, the gross enrollment rate (GER) stands at approximately 140%: this means that the expected number of enrolled primary pupils is 1,750,000, but the actual number is 2,450,000. In Secondary, NER and GER stand at 28.3% and 38.0% respectively. These patterns — of "bunching" in the early primary grade and progressively lower enrollment in higher grades — are depicted in Figure 2 below. The grey bars represent the baseline ideal enrollment, according to age, assuming no late enrollment, no repetition and no dropout. The colored bars represent the actual number of students enrolled.

[13] Glick and Sahn (2010)

Figure 2: Ideal (gray) vs. Actual Enrollment (color): "Bunching" in Early Grades[14]



The main drivers of these patterns are high repetition rates (RR) and high dropout rates (DR), as shown in Figure 3. While both rates are high, repetition is especially high in the earlier grades of primary, while dropout tends to increases towards the end of primary (as well as throughout secondary)[15]. While these trends hold across two different government data sets, there is uncertainty on exact national magnitudes of DR and RR figures. The first data source is "macro" school-level Ministry of Education (MINEDUC) data and the second is "micro", household-level national survey data (EICV). The EICV micro data has overall higher RR and lower DR[16].

Figure 3: Dropout and Repetition Rates from P1 to P6 in 2013[17]

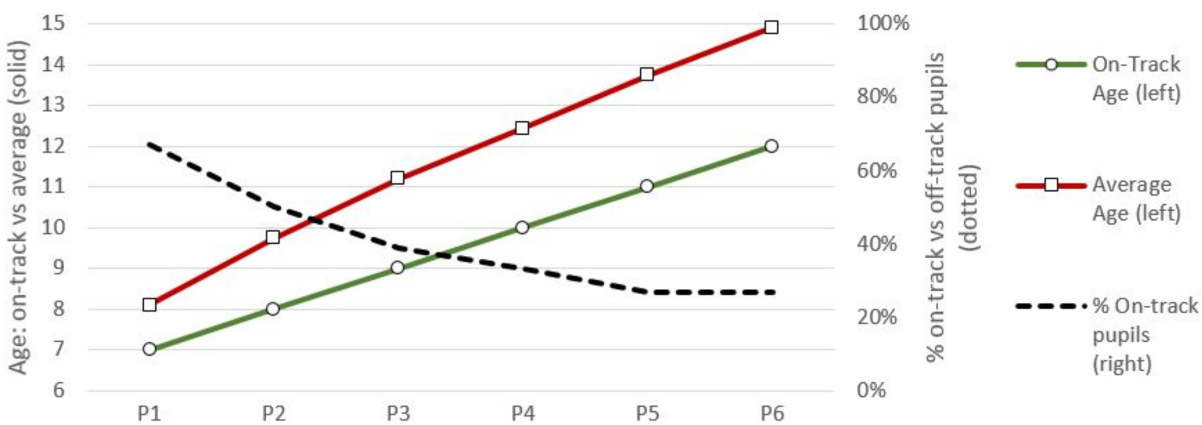
Primary Grade:		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
Repetition Rate:	MINEDUC (macro)	25.7%	15.9%	13.8%	14.2%	16.2%	2.7%
	EICV4 (micro)	*38.6%*	*23.8%*	*23.8%*	*16.9%*	*18.1%*	*7.0%*
Dropout Rate:	MINEDUC (macro)	10.2%	13.8%	12.7%	12.9%	28.3%	23.9%
	EICV4 (micro)	*1.6%*	*2.6%*	*4.3%*	*7.8%*	*11.9%*	*24.4%*

[14]Author's own work for Laterite's "*Mapping The Secondary Education System in Rwanda*". Not Published.
[15] The peak of RR and DR in P5 could be linked to a central exam in P6, on which schools want to have a large proportion of successful students. Schools may pre-emptively hold students back in P5 who they do not think are ready for the P6 exam. This artificially will improve the school's exam pass rate in P6.
[16] Micro data EICV based calculations could be the more accurate estimate as it is harder using aggregate macro school level data to take into account e.g. inter-school dynamics like "switching schools" in repetition and dropout calculations. This could explain the higher dropout estimates for MINEDUC macro data compared to EICV micro data: pupils changing schools show up as *dropouts* instead of *repeaters* or *promoters* in school records, as *school-changing-students* are indistinguishable from *dropout-students* from the school's perspective. On the other hand, EICV could be more prone to sampling error and fluctuations.
[17] Last year for which disaggregate data according to grade is available

Besides ‘bunching’ (See Figure 2 above), drastic over-aging also occurs. Figure 4 below illustrates the decreasing proportion of “on-track”[18] students, as well as over-aging. There is a growing gap between the “on-track” age and the “actual average age” of enrolled pupils. The gap is only 1 year on average for students in P1, but increases to an average of 3 years by P6. This suggests that overaging is due to repetition and not late enrollment.

While there has been a gradual decrease in dropout rates since 2007, there has been a recent surge in both repetition and dropout rate in 2012 (See Figure A1 in Appendix 10.1).

Figure 4: Increasing Over-aging Rate and Decreasing On-track Rate



3.3 Relevant Targets and Current Reforms in Education System

The 2013 to 2018 Education Sector Strategic Plan (ESSP) lays out MINEDUC targets for reducing both dropout and repetition as shown in Figure 5 below. These baseline numbers come from the MINEDUC Management Information System (MIS), the data source used for all government target setting. As is noted in the 2013 strategic plan, these figures estimate lower baseline rates of dropout and repetition than the national survey data (EICV).[19] While we use EICV data in our analysis section, due to greater granularity in the data, we do not make a judgement on which system-wide figures are more likely to be accurate.

[18] “On-track” here is defined as having the target age, plus or minus one year of the target (e.g. 6, 7, or 8 in P1 and 11, 12, or 13 in P6).

[19] “Education Sector Strategic Plan (ESSP) 2013/14 - 2017/18” p. 23

Regardless of data source, it is important to note that MINEDUC has made ambitious goals on reducing both dropout and repetition rates and has prioritized them on both educational and fiscal grounds. The 2013 ESSP, states:

> *Given limited financial resources to realize the ESSP, immediate attention will be given to improving the efficiency of the education system, by reducing repetition and dropout.*[20]
>
> *By improving the quality of instruction and availability of learning materials, particularly at the primary level, repetition rates are projected to fall over the course of the ESSP. This significantly reduces costs compared to paying for students to complete the same year of education twice and is a major economic justification for prioritising quality in the education sector.*[21]

Despite efforts to improve school quality since the 2013 ESSP, preliminary data for 2013 to 2015 show that repetition rates remain high (see Figure 5 below). This suggests that efforts to broadly improve educational quality may be insufficient in reducing repetition and a more targeted approach may be required.

Figure 5: MINEDUC Targets for Reducing Primary School Repetition and Dropout Rates[22]

Indicator (Source EMIS)	**2012/13 Baseline**	**2013/14**	**2014/15**	**2015/16**	**2016/17**	**2017/18**
Target: Repetition Rate	**12.7%**	11.7%	10.7%	9.7%	8.7%	7.7%
Actual: Repetition Rate		18.3%	20.7%	18.4%	N/A	N/A
Target: Dropout Rate	**10.9%**	9.3%	8.4%	7.6%	6.6%	5.7%
Actual: Dropout Rate		14.3%	10.3%	5.7%	N/A	N/A

On dropout, initial trends indicate that there has been improvement since the 2013 ESSP; however, additional data from 2016 and 2017 will be needed to confirm that the trend continues or stabilizes.

[20] Ibid p. 37
[21] Ibid p. 98
[22] "Education Sector Strategic Plan (ESSP) 2013/14 - 2017/18, p.87 and "2015 Education Statistical Yearbook." Ministry of Education, Republic of Rwanda, June 2016, p.27

4 Policy Context: National Development and Education Sector

This section explores key national policies relevant to dropout and repetition. The broader policy agenda in Rwanda provides opportunities to improve dropout and repetition rates through integration with existing programs. It also serves as a reminder that educational outcomes like dropout and repetition will affect broader policy outcomes in Rwanda's national development planning, particularly around the transformation of the economy. This section briefly summarizes key policy initiatives across the government more broadly, such as Vision 2020, and in the education sector, in particular, such as the introduction of Imihigo congrats and advances in ICT.

4.1 Vision 2020: National Development Strategy and Education

The Rwandan government has launched an ambitious industrial development plan called Vision 2020 that seeks to make Rwanda a middle-income country by the year 2020. This goal headlines a broader set of industrial policies that aim to transform the country and its economy from "a subsistence agricultural economy to a knowledge-based society."[23]

Vision 2020 acknowledges the critical role that the education system must play if Rwanda is to meet these ambitious goals. The industrial policy plan detailed in Vision 2020 emphasizes increased technical and vocational training among secondary and tertiary students. Improving this type of education it argues will be necessary to staff the sectors that the government is targeting for growth - logistics, telecommunications, and finance.

At present, the large number of students who regularly repeat early primary grades or dropout in primary school lead to low enrollment rates in secondary school and beyond. In order to ensure a large enough cohort of students ends up in technical and vocational school, it will be important to continue to work to improve dropout and repetition rates in primary grades.

4.2 Imihigo Contracts

Since its inception in 2006 the Imihigo performance contract system has been an important part of Rwanda's improvements in healthcare, infrastructure, and other aspects of the

[23] "Rwanda Vision 2020 (Revised 2012)." Rwandan Development Board, 2012. p. 8

country's developments.[24] In 2016, the Imihigo performance contract system was extended to both regular teachers and head teachers across the country.[25] At the teacher level, contracts currently layout goals along 11 dimensions, ranging from how prepared teachers are to success rate of students (full list available in Appendix 10.2.1). These dimensions do not include dropout or repetition goals.

However, head teachers have several additional dimensions upon which they can be evaluated, **including the reduction of dropout in their school**. Each teacher is evaluated against their annual Imihigo goals by their school's Head Teacher. Head teachers are in turn evaluated on their Imihigo contracts by Executive Secretary of their Sector.

Although structures have been put in place to build both repetition and dropout targets into Head Teacher's Imihigo contracts, to date this option has not been widely exercised.[26] In the policy analysis and recommendation sections below, we discuss how Imihigo contracts can be used to complement a broader strategy in reducing dropout and repetition.

4.3 ICT and Education: Better Data and Diagnostics

Improved information and communication technology (ICT) may help better target or deliver interventions to reduce repetition and dropout. In the 2016 Education ICT Policy, the Ministry of Education laid out an ambitious goal of 100% coverage for it's "Smart Classroom" policy by 2018. Along with generators and other infrastructure upgrades, the policy sets a goal of "One Digital Identity Per Child".[27] As a part of the initiative, the number of computers in the public school system is also set to drastically increase in the next few years.[28] Increased access to computers in classrooms may allow for digitized student testing and surveys that can be used to track student progress. There is also the possibility of using ICT technologies for remedial intervention to catch students up on basic skills. Recent studies in India have shown that this is a

[24] There have not been experimental evaluations of the Imihigo system. However, there has been some research on the effect of inclusion of specific goals in Imihigo contracts on service uptake, specifically family planning services, which showed large positive effects. See: Scher, Daniel. "The Promise of Imihigo: Decentralized Service Delivery in Rwanda, 2006-2010." Innovations for Successful Societies.
[25] "Determining Modalities for Performance Appraisal of Teachers." (2016)
[26] Ibid
[27] ICT in Education Policy. Ministry of Education (2016) p. 4, 9
[28] "Positivo, Made in Rwanda Computer to Increase Country's Export." (2015)

high potential use of ICT in the classroom.[29] Having laid out the problem and context, we now turn to a discussion of our analytic framework and methodology we use to develop solutions.

5 The Predictive Approach: Analytic Framework

The remainder of this policy brief discusses the opportunity and challenges of using "predictive analysis" to identify which students are most at risk of repeating grades. By prediction we mean two things. First, we focus on identifying **which students will repeat** rather than on identifying **what are the factors that cause** repetition. Second, prediction refers to a specific set of **"machine learning" analytical techniques** which are described in more detail below and in the technical appendix.

This section first explains why a predictive approach is well suited to this question and provides a brief introduction to machine learning. It then summarizes examples of these techniques in other policy sectors and geographies. It concludes with discussing the types of policies that best complement a predictive approach: targeted programming and incentive systems paired with individual discretion.

5.1 Why a Predictive Approach Makes Sense

Taking a "predictive approach" to the problem of student repetition of dropout has two key benefits: (1) helping better target scarce resources and (2) enabling preventative approaches.

5.1.1 Prediction Helps Target Scarce Resources

Although dropout and repetition rates are high in primary schools, in a given year more than 70%[30] of students will be promoted to the next grade - neither dropping out nor repeating. To most efficiently spend government resources, the ideal policies to prevent repetition and dropout would focus only on those identified as vulnerable. An accurate prediction algorithm provides a standardized way to focus resources on those students who need it most.

[29] Muralidharan et al. (2016)
[30] "2015 Education Statistical Yearbook", Ministry of Education, Republic of Rwanda, June 2016, p.27

5.1.2 "Preventative" Instead of "Curative" Policies

With an accurate prediction of which students are at the highest risk of dropout or repetition in the future, schools can intervene earlier. This is particularly important for problems like repetition and dropout where early intervention may be more successful and cost effective.

In the case of dropouts, once the student has left the school system they may be hard for a school to locate or hard to convince to come back to school. Finding each child that has dropped out and re-registering these children is a very expensive approach to addressing the issue at hand.

Likewise, when a child repeats a school year, they incur an entire school year of expenses to go through the curriculum again, even if they have repeated for reasons tied to a single subject. This both costs the school money and may limit the amount of learning that the child gains in that year. Earlier intervention has the potential to be cheaper and more successful.

5.2 Machine Learning: The Methodologies Behind Prediction Algorithms

With an increase in data availability and computing power, a new set of statistical techniques known as "Machine Learning" have become increasingly common in public policy and other domains. At its most basics, machine learning refers to a broad set of methods that work with large data sets and is used to generate and test predictions. The methods used in this paper are discussed in our analysis section and in the technical appendix (see Appendix 10.4).[31]

There are a few defining characteristics for the types of problems that machine learning is helpful in solving. All three of these characteristics apply to problems of dropout and repetition.

- **"Big" Data:** Machine learning approaches typically utilize large datasets numbering in the thousands of observations or ideally more.
- **Multifaceted Problems**: In addition to processing many observations, machine learning approaches are helpful in solving problems involving a large number of predictor variables and determining which variables have the highest predictive value.
- **Future Problems**: Machine learning approaches are also well suited to problems that involve predicting future outcomes based on similar past cases.

[31] For a more technical overview of machine learning in public policy settings see: Kleinberg et al. (2015)

5.3 Examples of Predictive Approach

Machine learning predictive analysis has been applied to a number of public policy issues across the developed and developing world in education and other public policy domains.

5.3.1 Predicting School Dropout and Other Outcomes in the Developed World

In the education field, machine learning has been used to predict which students in secondary school and university are most likely to dropout of school. Using a dataset of 72,000 Danish students, researchers in Denmark built a machine learning algorithm that could predict which secondary school students would dropout in the next 3 months with greater than 90% accuracy.[32] On a smaller scale, researchers in Mexico applied similar techniques with a considerably smaller dataset of 670 middle school students.[33] In 2011, the Bill and Melinda Gates Foundation funded a $1,000,000 grant funding the development of predictive analytics to combat student dropout across a number of different universities in the U.S.[34]

Predictive analytics have been applied to a variety of other public policy questions in the U.S. and other economically advanced countries. In cities, machine learning algorithms are used to predict where crimes are likely to occur to inform policing routes[35] and to predict which criminal suspects are flight risks.[36] There are numerous other applications in health care, such as predicting patient re-admittance,[37] and a variety of other social services.

5.3.2 Predictive Analytics in the Developing World

While less common in less developed countries, predictive analytics has been regularly used for a few specific use cases. The most common use case has been the development of Proxy Means Tests (PMTs) for the targeting of social services. In countries with large informal economies, exact income levels are difficult to calculate. Consumption surveys are expensive to calculate for even a nationally representative sample, let alone everyone in the country. As such,

[32] Sara et al. (2015)
[33] Marquez-Vera, et al. (2013)
[34] Vuorikar and Muñoz (2016) p. 132
[35] Perry et al. (2013)
[36] Kleinberg, Jon, et al. (2016)
[37] Lee (2012)

many countries use PMTs such as the Progress Out of Poverty Scorecard[38] to predict whether someone lives above or below a given poverty line based on easily observable variables. These scorecards are based on each country's own national survey data, in the case of Rwanda, EICV data is used. These types of predictions are used to determine eligibility for welfare programs or eligibility for services by NGOs.[39]

Machine learning approaches have also been used to estimate credit risk for borrowers who lack access to formal credit scores in the developing world. The Entrepreneurial Finance Lab (EFL)[40] uses psychometric surveys to predict whether poor borrowers are likely to default on loans. These services are used by a wide variety of banks 27 countries from Brazil to Zimbabwe.

5.4 Turning Predictions into Policy and Programs

Predicting repetition or dropout risk alone is not sufficient. Predictions alone only provide information on student risk, but do not provide a template on what comes next. There are two broad sets of approaches to public policy that compliment accurate predictions of future outcomes, such as dropout or repetition. They are 1) incentive systems with discretion and 2) targeted, pre-designed programs.

5.4.2 Discretion + Incentives

One approach would be to provide bureaucrats, in this case teachers, with the prediction results, set a goal of improving related outcomes, and then incentivize performance to reward accomplishment of these goals. This approach may make more sense if each student faces different types of challenges that must be addressed on a case-by-case basis.

5.4.1 Targeted, Pre-designed Programs

A second approach would use the predictions generated to implement targeted programs or interventions to the most at risk students. Such programs could target known causes of repetition or dropout, such as poor school performance or household features, such as inability to

[38] For background see: Progress Out of Poverty Index (www.progressoutofpoverty.org/about-us)
[39] For an example in research see: Alatas et al. (2012)
[40] For primary consulting service doing this work see: "How It Works." EFL Global, www.eflglobal.com/works/. (Disclaimer: SYPA Adviser Asim Khwaja was a co-founder of EFL)

pay supplemental school costs. This approach is likely to work best if most students at risk of dropout or repetition face similar underlying causes that can be addressed by a standardized program.

5.4.3 Program Design and Implementation

This report is agnostic as to which of these approaches is more likely to be successful, given the lack of data available to test these different approaches presently. As will be discussed in more detail in the implementation section (see Section 8), we recommend that MINEDUC explores piloting and evaluating both of the above options in combination with a predictive approach to identify at risk students.

These high-level design principles outlined in section 5.4 will be concretized in Section 7 with a consideration of the specific policy environment in Rwanda. Before detailing these policy proposals, we first show in section 6 that it is possible to build a predictive scorecard given constraints on data availability and resources.

6 Methodology

This section lays out the criteria for a successful predictive model, discusses the steps we took to build our model, and details how model accuracy should be evaluated. For our prediction approach to be successful, we want a model underlying the scorecard that is:

(1) **Accurate**. We want the time and resources spent on preventing repetition and dropout to go to those pupils who are *actually* "at-risk" (i.e. those who would have repeated or dropped out). There is an inherent tradeoff between improved accuracy and the following two goals.

(2) **Simple**. Ideally we want a scorecard that uses a simple and transparent model. First, this will increase trust and buy-in at all levels of the education system, from MINEDUC and REB officials to teachers in schools. Secondly, it will make implementation easier, as administering and creating a simple scorecard lowers the skills-requirements and relies less on specific expertise knowledge.

(3) **Limited data requirements**. Given that extensive and up-to-date data on pupils is not (yet) readily available, data collection will need to be part of the process. By keeping data collection simple and low cost, the feasibility of using scorecard predictions increases.

Our analysis shows that it is possible to build a predictive tool that satisfies these three goals. In this specific case, the tradeoffs between accuracy and cost are not as steep as anticipated as we will show below.

6.1 Data Used for Model Creation

The accuracy of prediction model will depend heavily on the availability and quality of data. While new, high quality is being collected[41], the best available data during this analysis are the Integrated Household Living Conditions Surveys[42] (EICV). Since 2000, EICV has been conducted every 3 to 5 years and includes individual data on demographics, health, education, housing, services, and economic data. For individuals of school going age, data on both the current grade of enrollment as well as previous year's grade is recorded, allowing us to determine whether that pupil promoted, repeated, or dropped out. These are the outcome variables we predict using most of the EICV4 data. Keeping only individuals enrolled in primary (the focus of our study), we are left with around 13,000 observations. After data cleaning, we end up with more than 150 variables from which to choose predictor variables for the eventual scorecard.

6.2 Variable Selection

We need to narrow down the number of variables used in the prediction for two reasons. First, not every variable is a good predictor of repetition or dropout and including poor predictors will make our model worse. Second, even among accurate predictors, it is important to identify which variables are the most predictive. By keeping only the most predictive variables, models can be kept simple and thus minimize data collection requirements.

[41] Laterite, as part of a "Dropout" study commissioned by the MINEDUC and UNICEF, is collecting data on 10,000 primary and secondary pupils from 3,000 households. This includes individual data (including the pupil's schooling history), household data, school level data, and community level data. This data will lend itself well to updating some of the analysis conducted here, including the scorecard itself.

[42] The National Institute of Statistics Rwanda (NISR) is Rwanda's central agency for data collection and dissemination. The EICV4 data was retrieved from: http://microdata.statistics.gov.rw/index.php/catalog/75

The quantitative technique we used for variable selection is Logistic[43] regression with LASSO[44] regularization. Adjusting the parameters of the LASSO model allow us to control the complexity of the model. This way can produce models that use only a handful of variables to predict the outcome, or we can produce a model that uses several dozen variables.

6.3 Which Predictor Variables to Choose: The 'Cost'/'Accuracy' Tradeoff

In prediction exercises, having more data is almost always better. However, **an increase in predictive power must be weighed against the additional cost of collecting more data**: a model with only a few variables will make data collection and scorecard creation easier and cheaper, but may result in less accurate predictions. A complex model with many predictor variables, on the other hand, might result in higher accuracy, but increase data collection and project management costs.

We find that there is only a small improvement in model performance when including more costly variables in our prediction model, which implies that a scorecard can be *low on data requirements*. We come to this conclusion when we assess the level of this trade-off by splitting the initial 150 variables into three categories, by difficulty and cost (see Appendix 10.3 for more details): from (1) easy-to-collect *'teacher'* variables at the school level, to (2) data collected at the *'community'* level, to (3) expensive *'professional'* variables that would require a survey team. Figure 6 below shows that including *'community'* or *'professional'* variables on top of the basic *'teacher'* level variables result in very low marginal gains in accuracy.

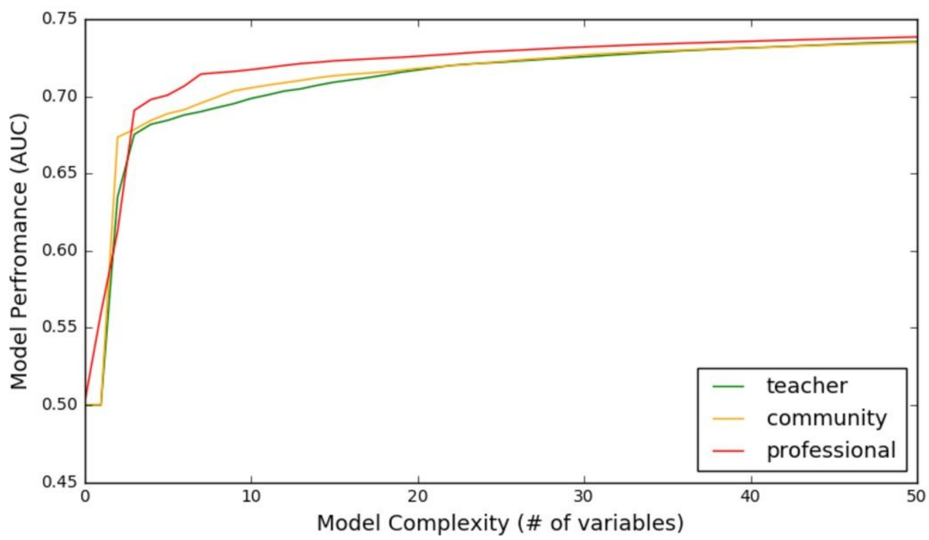
6.4 How Many Predictor Variables to Use: The 'Complexity'/'Accuracy' Tradeoff

Similar to the above tradeoff, **an increase in predictive power must be weighed against the inclusion of more variables.** More variables will increase accuracy, but will also make the scorecard longer and more complex to implement.

[43] Since the outcome of interest is binary (i.e. "repeated" or "not repeated"), a logistics model was out first choice. Moreover, using logistic regression for variable selection and for model creation outperformed other models like random forests. This is in line with other research predicting dropout (e.g. Aulck, L, et al. (2016))
[44] LASSO stands for "Least Absolute Shrinkage and Selection Operator". A more thorough overview of LASSO and our quantitative analysis can be found in Appendix 10.4.3.

Fortunately, **we find that increasing the number of predictor variables in our model beyond 10 to 20 variables leads to very low returns in increased accuracy, which implies that a scorecard can be kept *simple***. We reach this result by creating increasingly complex models, where 'complexity' means including more predictor variables. We do this for each of the three variable categories: *'teacher'*, *'community'*, and *'professional'* variables. The results are shown in Figure 6 below. On the bottom axis is the number of variables (predictors) used in each algorithm. As we move to the right, the scorecard would become more complex and more data would need to be to collected. The vertical axis measures model performance using the AUC metric: a higher value means a more accurate model. The AUC is discussed briefly in Section 6.6 below and more in-depth in Appendix 10.5.1.

Figure 6: After 20 Questions, Scorecard Accuracy Plateaus



Graph Note: The results of this analysis are promising, from a policy perspective, for two reasons. First, after a model exceeds approximately 20 or so predictor variables, adding more questions and data points does not drastically improve prediction quality. Second, the variables that are relatively inexpensive to collect, by a teacher or by a community member, perform nearly as well as the data that could be collected by a more experienced, but expensive, data collection team. Finally, note than 25 variables can be turned into a 10-12 question scorecard by condensing binary variables to categorical data.This is explained in detail in the technical appendix.

In light of these findings, our focus now turns to developing a simple 10 to 15 question scorecard, and how to best implement such a scorecard[45].

[45] We assume that the step from *'teacher'* level data collection to *'community'* is rather small, in terms of difference in implementation cost, especially compared to the step in going from *'community'* level to *'professional'* variables. Hence our choice to create a scorecard based on the first two.

6.5 Creating a Scorecard: From Regression Model to a Simple Scorecard

With a raw model in place, how do we get to a user-friendly, but accurate, scorecard? After the variable selection step (using LASSO, see above) applied to the community level data, we have model with the 25 most predictive variables. After combining multiple variables into single questions and a simplification process of the variable coefficients (see Appendix 10.5.2), we end up with a one-page scorecard with 11 questions. Each response corresponds to either 0 points or a round number of points towards a final score. The higher the score, the higher the repetition risk for the individual is. A proof-of-concept scorecard is given in the next page.

Note that:

- While the maximum possible score currently is 64, the points attributed to each question could easily be rescaled so that the total maximum score is 100 instead of 64.
- The resulting total score should not be interpreted as a direct and linear correspondence to a probability of repeating. Relative position is what matters: someone with a higher score is more likely to repeat than someone with a lower score.
- The AUC of this scorecard is 0.72. This figure, as described above describes the trade-off between ‘capture’ and ‘leakage’.

Score Card Draft I: Student Risk Scorecard
Community Data Collection

Student Name: ________________________________ Date: _________

Q1) Circle the cell based on the student's age and year

Grade | Age	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16+**
P1	8	8	7	6	5	4	3	3	2	1	0
P2	-	8	8	7	6	5	4	3	3	2	1
P3	-	-	12	12	11	10	9	8	7	7	6
P4	-	-	-	15	14	13	12	11	10	9	9
P5	-	-	-	-	18	17	16	15	15	14	13
P6	-	-	-	-	-	10	10	9	8	7	6

Score from table: ______

Questions for Parents	Circle Response	Points	Score:
Q2) Does the student own a mobile phone?	Yes No	0 7	______
Q3) Has the student fetched water in the last 7 days?	Yes No	3 0	______
Q5) If student has missed school in last 7 days, was it for any of the following reasons?	Poor Health Family Circumstances Hasn't Missed School	4 6 0	______
Q6) Has the child worked for at least 1 hour in the last 7 days?	Yes No	0 1	______
Q7) Has the child worked in non-farm business for no pay in last 7 days?	Yes No	5 0	______
Q8) Can the student **read** a letter / simple note?	Yes No	0 8	______
Q9) Can the student **write** a letter / simple note?	Yes No	0 7	______
Q10) Can the student perform a written calculation?	Yes No	0 3	______
Q11) Did the family pay any school tuition this year?	Yes No	0 6	______

Final Score: _________

Community Surveyor Completing Scorecard: ______________________ Date: _________

6.6 Prediction Accuracy: Measuring Scorecard Performance

6.6.1 General Model Performance

How do we assess how well a model (and eventually a scorecard) performs in predicting repetition (or dropout)? There are two important goals to consider:

(1) Minimizing "undercoverage" or Maximizing "Capture": Before a student actually repeats, a good scorecard would label the pupils that eventually **will repeat / dropout** as being at-risk. That is, we don't want 'miss' any actually at-risk pupils by failing to predict them as being at-risk. In Figure 7 below, we want to minimize #2 and maximize #1.

(2) Minimizing "leakage": We do not want to identify pupils as at-risk if they would not have repeated in the first place. We want a scorecard that minimizes the prediction of "false-positives", that is, at-risk for pupils that are in fact not at-risk. Resources invested in these pupils is not strictly needed. The student would have passed anyway. In Figure 7 below we want to minimize classification in #3 and maximize #4.

Inevitably, *any* model or scorecard will have to make a trade-off between these two: higher "capture" will lead to higher "leakage" (i.e. when a scorecards predicts relatively more pupils as at-risk, left column A), but lowering "leakage" will lead to lower "capture" and increased "undercoverage" (i.e. more are categorized as not at-risk, in right column B).

Figure 7- Goals of the prediction	A) Predicted as at-risk	B) Predicted as **not** at-risk
Actually at-risk (i.e. will repeat)	*#1: Maximize ✓ ("Capture")*[46]	*#2: Minimize ✗ ("undercoverage")*
Actually **not** at-risk (i.e. will **not** repeat)	*#3: Minimize ✗ ("leakage")*	*#4: Maximize ✓ ("correct not-at-risk pred.")*[47]

[46] This is also technically referred to as *Sensitivity*
[47] This is also technically referred to as *Specificity*

The main performance indicator used is the AUC value (e.g. see Figure 6). The AUC captures the above trade-off in a single number (see Appendix 10.5.1 for more on AUC). A perfect prediction model has an AUC of 1: *all* at-risk pupils are correctly predicted as such (perfect "Capture"), and *none* of the not-at-risk pupils are predicted as being at-risk (no "leakage")[48]. A 'model' that would randomly 'guess' who is at-risk would have an AUC of 0.5. Hence, the further away from 0.5 and the closer to 1 the AUC value is, the better the model.

AUC can be used to compare one model to another. However, once a final model has been selected, we must also choose a cut-off score in our model. All students above the cutoff score would be flagged as "high-risk" and all students below the cutoff score would be flagged as "lower-risk". This could be done in multiple different ways. We could select, for example, the 10% of students with the highest scores. Or we could pick a cut-off score that is used across schools. As will be shown below, there is a tradeoff between leakage and undercoverage when setting such a cutoff.

6.6.2 Draft Scorecard Performance

To get an idea of the spread of scores (for the 12,500 pupils in our sample) we plot the frequency of the students' scorecard scores in a histogram. This is shown in the top panel of Figure 8 below. Very few pupils have a score below 10 or above 50 (i.e. out of max 64). The blue bars are the students who do not repeat while the green bars represent the repeaters[49]. In line with expectations, almost half of the pupils with a scorecard score over 40 are repeaters. Of those pupils with scores below 30, only very few actually repeat (i.e. very little 'green').

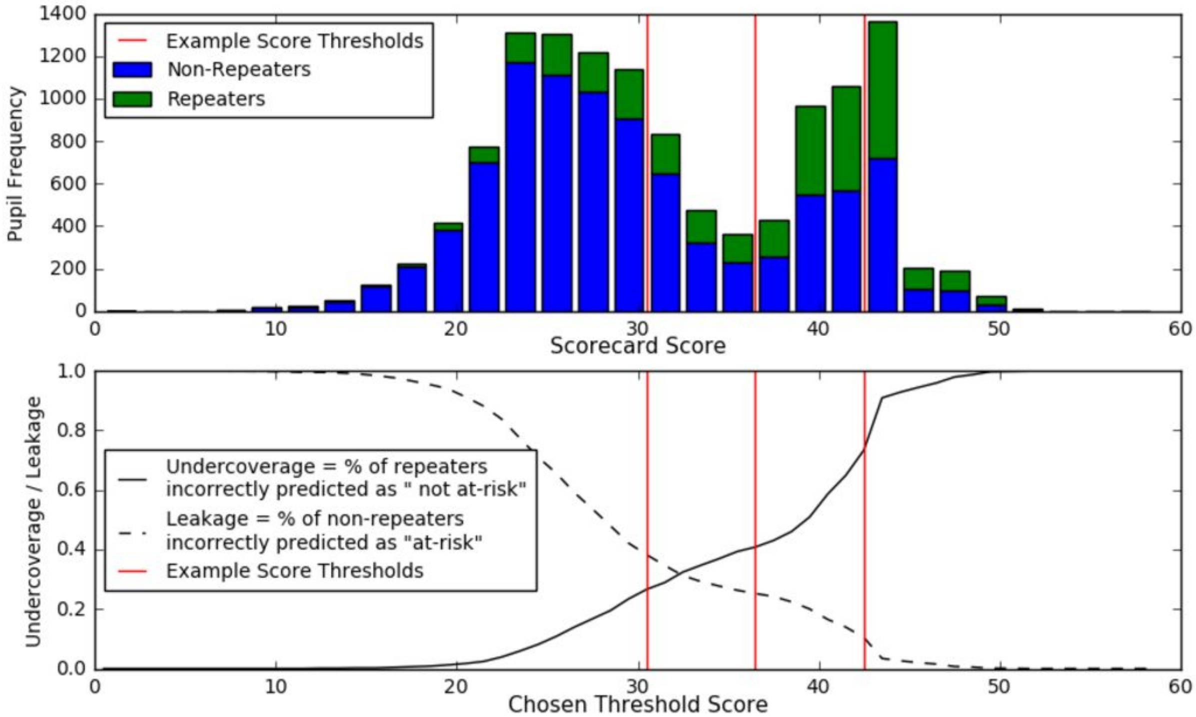
The lower panel shows the tradeoff being made when choosing a threshold. A *higher* threshold for being classified as "at-risk" means:

1) *lower 'leakage'* (i.e. the less non-repeating students we falsely classify as being at-risk),
2) but also *higher 'undercoverage'* (i.e. we fail to identify more of the repeaters as at-risk).

At lower thresholds, we 'capture' more repeaters (i.e. lower undercoverage), but we also misclassify more non-repeaters as "at-risk".

[48] This is the situation depicted in the upper left panel of Figure 9 below
[49] Note that the disaggregation into blue (non-repeaters) and green (repeaters) is ***not*** info we will have in the future.

Figure 8: Distributions of Scorecard Scores and the Leakage-Undercoverage Trade-off



Different choices of threshold will have different implications. For example, what do 3 different threshold values (also indicated in red in Figure 8 above) imply for a hypothetical class of 100 pupils, 26 or which who will actually repeat if there is no intervention, while 74 who will not.[50] Figure 9 below summarizes what the classification results would be for each of the three thresholds.

While a threshold of 31 might be too low (i.e. we classify almost half of the pupils as at-risk, but only 60% of these would actually repeat), a threshold of 43 is probably too high (i.e. very few of the actually at-risk pupils are 'captured'), and "the best" threshold will probably be somewhere in between. However, there is a clear trade-off and choosing "the best" threshold will depend on financial considerations and other stakeholder priorities.

[50] Note that for clarity's sake, we leave aside the students who dropout, which in the EICV are only 5% of the observations.

Figure 9: Three Example Thresholds in a Representative Classroom of 100 pupils

*The impossible **perfect** model*	Predicted as at-risk	Predicted as **not** at-risk
Actual repeater	**26**	**0**
Actual **non**-repeater	**0**	**74**

*Scorecard with threshold of **31+***	Predicted as at-risk	Predicted as **not** at-risk
Actual repeater	**19**	**7**
Actual **non**-repeater	**28**	**46**

*Scorecard with threshold of **37+***	Predicted as at-risk	Predicted as **not** at-risk
Actual repeater	**15**	**11**
Actual **non**-repeater	**19**	**55**

*Scorecard with threshold of **43+***	Predicted as at-risk	Predicted as **not** at-risk
Actual repeater	**7**	**19**
Actual **non**-repeater	**8**	**66**

Figure 9 Notes: For a threshold of 31+, we would only 'miss' 7 out of 26 actual repeaters, but incorrectly predict 28 out of 74 pupils as "at-risk". For a higher threshold of 37+, less of the repeaters are captured (only 15 out of 26), but we also reduce the undesirable 'leakage' from 28 to 19 pupils. At a very high threshold of 43+, 'leakage' is very low (only 8 non-repeaters as "at-risk" predictions), but we 'miss' 19 out of 26 pupils who would have repeated.

6.7 Limitations and Other Considerations

6.7.1 Limitations of the EICV Data

The quality of the data will be a limiting factor in how well a prediction model performs on new, *out-of-sample* and *out-of-time* data. In EICV4, data is collected in 2013/14 (i.e. prediction variables are for that year), but repetition and dropout data is about transitioning from 2012/13 to 2013/2014. Hence, we are using 2013/14 data to 'predict' events in the past. For children repeating, the 2013/14 data is probably a good representation of their 2012/13 situation, as they are still in school and the individual and household situation will be mostly similar. It is unclear whether this aspect of the data underestimates or overestimates the model's ability to correctly predict pupils' repetition risk.

However, pupils that dropout see drastic changes to their lives. Many start working or will get more engaged in other household activities. The 2013/14 data is *not* a good representation of their 2012/13 situation, and a model predicting dropout is overly optimistic about the ability to correctly predict dropout[51].

6.7.2 Model Consistency Across Time

While the prediction model creation process involves steps to prevent the model being too specific to the current data and timeframe of data collection[52], the model inevitably picks up on both sample-specific quirks as well as time specific elements. The factors that are very correlated with repetition and predict it well in 2013/14 might not be the ones that predict best in later years.

Therefore, the prediction model underlying a scorecard should be regularly updated: at least for every new EICV version, but ideally on a yearly basis. The feasibility of yearly updates will depend on the future of data collection. An expanded and regularly updated EMIS[53] with both individual, school, and possibly household level data would form the ideal basis to create or update prediction models on. With smartphones becoming more ubiquitous and affordable, these offer a promising platform for data collection, and some programs in Rwanda already have teachers collect student level data via smartphones[54].

Moreover, the longer data is collected and the more years of data are available,[55] ideally data on the same individuals, the more accurate a prediction model will be. In that sense, the scorecard presented here would be a baseline, lower-limit performing implementation, and future scorecards will improve upon it.

---

[51] The model mostly chooses work-related variables as predictors, which changed the most. See Appendix 10.3.2.
[52] We split the existing data into training and testing data, e.g. 70-30% split. See Appendix 10.4 for details.
[53] Education Management Information System
[54] For example, Right to Play has given teachers in Nyarungenge District 114 smartphones to "*give reports to education officials concerning the progress of play-based child learning activities, among other issues, in their respective schools. [...] The phones contain a software application in which the teachers will feed the necessary information [...]. Previously the teachers were using pen and paper to do this.*" (Kuteesa (2017), New Times Rwanda). Similar innovations in data collection on a wider scale would allow more frequent and accurate prediction model creation and automate student level predictions.
[55] i.e. a growing panel data set: the same variables collected for the same individuals over several years

7 Recommended Policy Options

In this section, we propose two separate policy approaches that incorporate predictive scorecards - (1) "**Incentive**" based policy and (2) "**Program**" based policy. A summary of these two proposals appears in Figure 10 below. **We are agnostic as to which of these two options is preferable given the information available today and recommend the Ministry run an experiment testing different approaches against each other**.

Figure 10: Policy Option Proposal Summary

Option 1: Incentive Based Policy Predictive scorecards identifying pupils at risk of repetition or dropout could help teachers prioritize which students to focus on. Imihigo targets should be set (1) at the school level for dropout and (2) at student level for repetition.
Option 2: Program Based Policy A specific program is designed to combat dropout or repetition (Ex. remedial after school lessons) by MINEDUC with its partner organizations. Predictive scorecards would then guide which students *are* or *are not* selected to participate in this new program. If relevant to the designed program, Imihigo provisions around dropout and repetition could be added for those actors involved in implementing the new program.

We believe both options are particularly well suited to the Rwandan context. Rwanda places a large emphasis on "homegrown solutions" or "Rwandan solutions for Rwandan problems". Approach 1, "Incentive Based Policy", builds on the Imihigo system of performance contracts unique to Rwanda. Both approaches 1 and 2 would be at the international frontier of education policy, building on emerging practices from industrialized education systems. Finally, both approaches attempt to reduce dropout and repetition without sacrificing on education quality for all students, a key goal of the government. The rest of Section 7 explores the potential pros and cons of each policy option. Implementation considerations are explored in more detail in Section 8 below.

7.1 Incentivized Discretionary Action: Scorecards & Imihigo

Policy option 1 embraces the view that predictive scorecards could be used to *guide* discretionary, individual action by teachers and other school staff, while Imihigo contract provisions would *incentivize* such action. Under this policy option, Teachers would receive basic training on how scorecards work as well as a list of students deemed "at-risk" by the prediction model. Data collection would occur at community meetings.

There are several possibilities as to how Imihigo could be structured. **We recommend that Imihigo contract goals focused on improving repetition are set at the teacher level and goals focused on improving dropout are set at the school level.**

Contracts could be set at the teacher level, head teacher (school) level, sector, or district level or even at multiple levels. Imihigo contracts set at a "lower level" - in this case the teacher level - may provide stronger incentives in pushing teachers to directly address the problems of students in their class. However, teachers may be discouraged if they perceive that they have little ability to improve the outcome by which the direct incentives are evaluated. In this case, teachers may perceive of find that they are unable to prevent student dropout themselves.

On the other hand, contracts set at a "higher level" - in this case at the school or sector level - may push more powerful actors such as sector officials or head teachers to direct resources to address the problem. However, these higher-level incentives may fail to motivate lower level teachers to exert maximal effort to prevent dropout or repetition because they do not have a direct stake in the outcome.[56]

Although closely related, dropout and repetition as policy problems have different characteristics. This suggests that two different incentive structures are required. Student repetition is more directly related to education in the classroom than student dropout, suggesting that teachers can better directly address repetition risk related problems. As such, we expect that applying incentives to improve student repetition directly to teachers can maximize teacher effort.

[56] This analysis draws from the broad study of *Contract Theory* in microeconomics. For more information see Gibbons (2013)

Alternatively, solutions to reducing dropout are likely to be more multifaceted, including intervention in the classroom and potentially at the household level. Individual teachers may be hard pressed to reduce dropouts by themselves. However, a school as a whole may be able to address these multiple root causes: some teachers focus on remediation and other teachers or staff focus on the community factors associated with dropout. This analysis suggests that dropout-reducing incentives are better suited at the school or head teacher level, ideally with all teachers receiving reward for meeting school wide goals. With this incentive design in mind, we now consider the positives and negative aspects of such an approach.

Option 1 - Problems this Policy May Solve

This policy option helps solve four potential problems that may currently lead to dropout and / or repetition. These issues would not be solved by an Imihigo contract (with repetition or dropout targets) alone.

- First, teachers may not prioritize students at risk of dropout or repetition because they are not rewarded for doing so. Adjusting Imihigo contracts to contain specific goals on repetition and dropout may help improve the situation, regardless of scorecard use.
- Second, teachers may lack information as to which students are at the highest risk of dropping out or repeating, particularly in schools with large class sizes and double shifts. Providing this information in tandem with the incentives in the Imihigo contract may be sufficient to improve repetition and dropout rates.
- Third, even if teachers have an accurate idea of which students are at the highest risk, there may be cognitive barriers that prevent teachers from focusing on these students. Teachers face a large number of simultaneous demands on their time in the classroom. They may not have the mental "bandwidth" to consciously prioritize which students are most at risk.[57] By providing a list of "at-risk" students to a teacher, the scorecard makes focusing on at-risk students more salient and actionable.
- Fourth, if students drop-out or repeat for a variety of different reasons, an individualized approach will be most effective in improving outcomes. Scorecards would identify at-risk students, while leaving room for teachers and schools to build personalized solutions to

[57] This argument come from the idea of "bandwidth scarcity", see: Mullainathan and Shafir (2014)

improve outcomes. In a context of heterogenous root causes, this could be a more effective approach than a "one-size-fits-all" programmatic solution.

Threats to Success Specific to Option 1

Option 1 emphasizes tailored, individualized attention by teachers and staff to specific at-risk children. In a resource constrained environment, both in terms of resources and cognitive bandwidth, providing individualized services to address repetition and / or dropout risk may crowd out time spent on other important aspects of a teacher's job. Alternatively, due to time constraints, teachers may simply ignore the risk predictions from scorecards and continue with their normal practices. This may be particularly true if incentives for improvement are too weak or if teachers believe they are not able to improve outcomes with the resources they have. More threats to success, applicable to both options, are discussed in the next section.

7.2 Better Targeting and Resource Use: Scorecards & Remedial Programs

Policy option 2 embraces the view that predictive scorecards could be used to better target new or existing programs addressed at students at high-risk or dropout or repetition. Like option 1, scorecards would be implemented in community meetings. However, unlike option 1, students identified as "high-risk" would be flagged for participation in a remedial program (rather than flagged for follow-up by teachers who would use discretionary action).

This program would be best developed by sector specialists at the Ministry of Education in consultation with partners such as UNICEF and Laterite. Laterite is currently conducting a nationwide survey assessing the root causes of dropout and repetition for MINEDUC. Based on the results of that survey and analysis, a more tailored program can be designed. Given that this detailed research is forthcoming, we will remain agnostic on program design until those results are available.

After program design, if relevant and appropriate to the specifics of the program, dropout and / or repetition goals could be added to the Imihigo contract of those civil servants responsible for implementing the new program. Imihigo contracts could also be tailored to include not only final outcome variables like repetition, but also include achievement of process outcomes, such as implementing the prescribed repetition- and dropout-mitigating programs well.

Option 2 - Problems this Policy May Solve

Similar to option 1, policy approach 2 would provide information, focus, and general economic incentives to solving the problems of repetition and dropout. As a tangible program rather than an abstract goal, option 2 may be more likely to focus school attention and limited resources on the problems of dropout and / or repetition. This program may be more effective than option 1 if teachers are unsure of what concrete steps they can take to address the problem.

Moreover, this option is more likely to be successful if students who drop out or repeat mainly do so for the same reasons (i.e. homogeneous underlying causes), allowing for "one-size-fits-all" solutions.

Threats to Success Specific to Option 2

Implementing option 2 would be more expensive than option 1. It would require the implementation of a new program which will come with extra human resources, training, and administrative costs.

Additionally, because this policy option outlines a specific program to be implemented, the "stakes" of a student appearing on a predicted at-risk list are higher. To potentially further improve correct "at-risk" identification and prediction performance – as well as to generate greater teachers buy-in – teachers could be given the option of changing the students in the pre-selected lists with other students in their class, up to a certain quota (e.g. 10%, 3 students, etc.).

Finally, as opposed to option 1, option 2 is likely to fail if the causes of student dropout and repetition vary drastically from one student to the next. If students all require individual attention and tailored solutions, designing one program for all at-risk students is unlikely to help.

Generic Threats to Success for Both Policy Options

There are three potential challenges that apply to both of these policy proposals that must be taken into account to maximize chances of success: (1) administrative challenges around scorecards, (2) lack of trust in predictions, and (3) potential self-fulfilling prophecy of labelling students as "at-risk". We believe all of these challenges can be addressed adequately by careful implementation summarized in Section 8.

First, there will be challenges administering the scorecard surveys. We propose administering these surveys at community Umuganda meetings to survey both parents and children. This will provide the opportunity to interview parents while excluding a child's specific teacher to reduce bias in survey results. However, there will likely be parents or guardians who are unable to attend meetings. There may also be challenges in matching parents to students and students to school records if the data is not collected at the school.

Second, even if accurate data is collected and predictions are calculated, there is a risk that teachers may not *trust* the results of the scorecard. A growing body of research suggests that humans often do not trust algorithms, instead preferring human judgement even when it yields inferior results. Moreover, humans tend to punish algorithms more harshly for mistakes than they do decisions made by humans.[58]

Third, there may be some risk in labelling students as highly likely to dropout or repeat. Research from schools in developed countries has shown that expectations about student performance can turn into self-fulfilling prophecies, referred to as the Pygmalion effect.[59] Simply telling a teacher that a student has exceptional potential can raise that student's performance. Likewise, teacher expectations of low performance or low ability can by itself lead to worse educational outcomes. Solutions to these challenges will be considered in the next section.

8 Implementation Considerations - Recommended Approach

8.1 Implementing Scorecard

This section outlines a series of recommendations on who should be responsible for implementing a scorecard, where it should be implemented, and how. We recommend that:

- scorecards be administered by a school official or community member at the monthly Umuganda community meetings across the country;
- data be collected twice per year, during the 1st and 2nd (of 3) school terms.
- information be collected using a paper-based template at first, but eventually migrate to an Android Smartphone based data collection application.

[58] Frick (2015)
[59] Rosenthal and Jacobson (1968)

The following sections expound on the reasoning for each of these recommendations.

8.1.1 Who Does Data Collection

Data for scorecards can be collected by several different individuals, including teachers in schools, members of Umuganda councils, appointed community representatives, SEOs (Sector Education Officers), or others. Since SEOs' responsibilities already pertain to the collection and aggregation of education data in their sector and schools, they can be responsible for coordinating implementation of this community-based data collection process. Regional Inspectors (RIs), as per their mandate, would be tasked with monitor adherence to proper program implementation, including data collection. (See Appendix 10.6 for an education sector stakeholder overview and interactions mapping.)

At community meetings, data can be collected directly from parents, which may be more accurate than interviewing students themselves. Collecting data at community meetings has additional advantages. The public nature of Umuganda meetings offers an opportunity to discuss issues of repetition, dropout, and other important school information with parents and community members. Moreover, community meeting surveys may increase transparency, generate more buy-in, and improve data quality.

Alternatively, teachers could complete the survey in their classrooms with each student. This approach is logistically easier than collecting data at community meetings, but has several drawbacks. First, busy teachers may not diligently collect data, or tweak survey responses based on impressions of certain children. Second, in schools with many classes, this data collection process may significantly detract from classroom time.

8.1.2 How Data Should be Collected: Technology and List Generation

In our draft scorecard we have presented a simple scorecard that can be used to administer the scorecard questions and score a child's risk of repetition that year. However, collecting data digitally has several significant advantages over collecting data using paper scorecards.

First, a digital approach can make data collection significantly easier for the surveyor. Using a mobile application, a scorecard survey can be programmed so the surveyor only has to input the answer to the questions on the scorecard. The application would then calculate each student's score and save the results in a database. This will eliminate the math errors that human enumerators are prone to make with pen and paper. It also opens up options for more advanced and accurate prediction models, not easily implementable using an additive scoring model.

Additionally, using an application will "hide" the point values assigned to each question. Displaying the point values may bias the answers recorded by an enumerator who wants to increase or decrease a child's score based on their own perceptions. This may also reduce the chance of those administering the scorecard thinking of the survey questions as causal factors. The prediction variables do not necessarily have a causal interpretation. Addressing these factors in a child's life will most likely not change the *actual* risk, only the calculated risk, which would worsen the prediction accuracy. For example: the act of a child fetching water may not in and of itself ***cause*** a student to repeat grades or dropout of school; however, it may just be ***correlated*** with those outcomes.

Second, digital data collection allows for better data storage, analysis, and reuse. Collecting student data year on year allows for the building of a rich dataset on student characteristics that can be used to improve future predictions. It will also allow for simple creation of lists of students who are most at risk (e.g. by class, by school, by region) and easy sharing between relevant school, sector, and district officials.

While a digital solution is preferred, the perfect should not be the enemy of the good. Piloting a scorecard on paper to familiarize actors in the school systems with the idea behind a scorecard is a sensible first step. If this is implemented, teachers should not be used to survey their own students to the reduce the risk of biasing results. If due to staff constraints, teachers do end up collecting the data using pen and paper surveys, the point values for each question should be removed to limit the bias effect. The administrative staff could calculate the risk-score.

Over the long-run a scorecard approach could be built into the broader EMIS[60] system. As the education system continues to improve access to ICT, it will become increasingly possible to conduct data entry into EMIS directly at the school level.[61] This will allow schools to

[60] Education Management Information System
[61] ESSP (2010) p. 57

collect student level data rather than school level data. In the long-run, predictive analytics for dropout and repetition may be run exclusively on data from the database.

8.1.3 How Frequently to Administer Scorecards

Scorecard surveys should be administered no more than twice per school year - once at the beginning of the school year and once after the first or second trimester concludes. The first time the survey is implemented, the full survey should be completed, including a basic learning level diagnostic (something like the ASER exam)[62] and household characteristics. If and when the survey is completed a second time in the school year, only those questions that are likely to have changed (such as learning levels, household status questions, etc.) should be asked and updated. For questions related to learning levels, schools' end of trimester exams could be used to ask key questions. Again, combining new and old survey data will be easier if data is collected digitally.

8.1.4 Overcoming Scorecard Risks

Section 7 identifies three key risks related to scorecard implementation: 1) data implementation difficulties, 2) lack of trust in the scorecard, and 3) adverse effects for selected children. We propose the below high level guidance to address these potential concerns.

(1) Implementation difficulties are best addressed by early and regular piloting and testing of the scorecard by those tasked with implementation of scorecards.

(2) To combat potential skepticism, the predictions can be framed as guidelines, where teachers have the ability to add or remove some students from the at-risk list, based on discretionary judgement.[63]Another way to improve trust in algorithms is to conduct a demonstration case. Teachers would guess at the beginning of the year which students are most likely to drop-out or repeat and at the end of the year their guesses are compared to the algorithm.

[62] The ASER test is a simple diagnostic tool used to assess basic literacy and numeracy, developed in India. It is available in English and could be easily translated into Kinyarwanda and adapted to the Rwandan context. Examples available at: www.asercentre.org/p/50.html

[63] This may increase algorithm accuracy if teachers are well informed. Alternatively, it could decrease predictive accuracy if teachers are not well-informed. In this case, a balance must be struck between trust-building measures and loss of predictive accuracy by changing predictions.

(3) Finally, labelling students as “high-risk” could turn into a self-fulfilling prophecy. By tying teacher level incentives to performance outcomes, the broader policy should counteract problems of labelling students. To more explicitly in address this problem with incentives, performance bonuses could be paid to teachers or schools who achieve high pass rates or non-dropout rates specifically among those students labeled at risk. Despite these safeguards, it will be important to monitor adverse effects during any pilot and experimentation phases.

8.1.5 Scorecard Cost Tradeoffs

Scorecard implementation costs can be minimized by building off existing resources and institutions. By using existing community meetings as a staging ground and relying on pen and paper implementation, there will be few direct costs in implementing the scorecards. The primary drivers of cost will be (1) initial training sessions on how the scorecards system works, and (2) staff time in entering and processing data (also an in-kind cost). If implementation of the scorecard is done digitally, costs in hardware and programming will go up, but overall teacher implementation time will go down over the long run.

8.1.6 Scorecard Implementation Wrap-Up: Starting with a Pilot

A small pilot will be necessary to finalize the approach of implementing the scorecard. There are likely to be both practical constraints not yet considered as well as unforeseen potential innovations once the scorecard is tested in person by those who would be implementing it long term. Establishing firm and feasible processes will be important before scale can be achieved.

8.2 Implementing Imihigo

Section 7 discusses in detail various options for designing Imihigo contract structures. Here we briefly consider other implementation considerations related to target setting and cost. The most difficult remaining design choice will revolve around setting target thresholds for decreasing repetition at the teacher level and dropout at the school level. Various levels of goals should be tested in the pilot phase to understand what is feasible for a given teacher or school to achieve in a school year. Goals should be attainable, but not too easy to meet, representing substantial progress if they are indeed met.

The cost of implementing Imihigo provisions for dropout and repetition should not extend past the costs of implementing Imihigo contracts more generally. Payout of salary bonuses is tied to the percent of goals reached in a contract rather than to each item in the contract. This means there is no "marginal cost" in adding dropout or repetition goals as long as the payout structure remains the same.

8.3 Implementing Design Program

As discussed previously, this report remains agnostic on what specific programs will best compliment predictive scorecards. Below we outline a high-level proposal on how to select a program to test and identify likely characteristics of a successful program.

Programs should address underlying causes and drivers of repetition and dropout. For repetition, these drivers are more likely to be school and learning related, hence programmatic interventions could be school and learning based. Drivers of dropout, while partially learning related, are also likely to have root causes in household and other socio-economic factors, so programs aimed at reducing dropout could be more community based. Much research on these drivers has been conducted and some existing literature might apply to the Rwandan context:

- One area is returns to education. At an individual and household level, the perceived returns might be lower than the actual returns (which are high in Rwanda[64]). This would lead to individual underinvestment in education[65]. In the past, programs providing information to children and parents have proven to be successful[66]. Higher awareness about the value of schooling could trigger more effort to stay in school and progress to higher grades.
- At the school level, an overambitious curriculum is also often a barrier to learning and progression[67]. Both of these can and are being addressed in Rwanda by transitioning to a more labor market relevant and competency-based 12YBE curriculum[68]. Adapting the content and speed of curriculum to pupils at risk of repeating could help them progress.

[64] Montenegro and Patrinos (2014)
[65] Chari and Maertens (2014)
[66] Jensen (2010)
[67] Pritchett and Beatty (2012)
[68] Education Sector Strategic Plan 2013/14 – 2017/18

- Idiosyncratic and area-wide economic shocks can also lead to dropout[69], in which case programs could help at-risk children's families with, for example, schooling related costs.
- While Rwanda has fee-free public schooling, evidence from Rwanda[70] and Uganda[71] shows that informal schooling costs could still be a barrier, in which case a program could focus on schooling cost subsidies or targeted cash transfer to at-risk children[72].

Laterite is working with MINEDUC and UNICEF on data collection and analysis that will help identify what some of those causal mechanism and drivers of repetition and dropout are in the Rwandan context. Both existing and this new research can further inform program design.

Good programs will address some of these root causes and will also fit into the "preventative framework" introduced in this policy proposal. Programs that involve getting students who are already out of school to re-enroll, while important efforts, are not good compliments to this approach. As the use of predictive scorecards in this way is novel given the context, it is very likely that external funding can be found to fund any pilots of such programs.

8.4 Implementing an Experiment

Our final implementation recommendation offers considerations for running an impact evaluation to best understand the relative merits of the two approaches we have laid out. Running a randomized controlled trial, randomized at the school level, would give precise estimates on the expected impact of scaling up some of the ideas proposed in this report nationally.

The final design of repetition or dropout programs will dictate many parameters of any experiment, including cost, number of schools in the experiment, and more. Leaving these aside, the ideal research should be designed with the following goals in mind: (1) to accurately measure the impact of on key outcomes of dropout and repetition, (2) measure any negative effects on student learning or in other areas of school performance caused by these interventions, and (3) understand the implementation challenges incumbent in rolling out any successful interventions at scale. We hope that such an experiment will provide MINEDUC with compelling evidence to make progress on this important policy issue of student dropout and repetition.

[69] Woldehanna & Hagos (2015)
[70] Williams et al. (2015)
[71] Deininger (2003)
[72] Lincove (2012)

9 Bibliography

"About the PPI." *Progress Out of Poverty Index*, www.progressoutofpoverty.org/about-us.

Alatas, Vivi, et al. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*, vol. 102, no. 4, June 2012, pp. 1206–1240.

ASER Centre. www.asercentre.org/p/50.html. Accessed 9 Mar. 2017.

Aulck, L, et al. *Predicting Student Dropout in Higher Education*. International Conference on Machine Learning (ICML), 2016.

Brophy, Jere. "Grade Repetition." *International Academy of Education*, vol. 6, 2006.

Chari, A.v., and Annemie Maertens. "Gender, Productive Ability and the Perceived Returns to Education: Evidence from Rural India." *Economics Letters*, vol. 122, no. 2, 2014, pp. 253–257.

Chihana, Menard, et al. "Failing to Progress or Progressing to Fail? Age-for-Grade Heterogeneity and Grade Repetition in Primary Schools in Karonga District, Northern Malawi." *International Journal of Educational Development*, vol. 52, Jan. 2017, pp. 68–80.

Deininger, Klaus. "Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda." *Economics of Education Review*, vol. 22, no. 3, 2003, pp. 291–305.

"Determining Modalities for Performance Appraisal of Teachers." *Instructions of Minister of State, Rwanda*, 4 Oct. 2016.

" Education Sector Strategic Plan (ESSP) 2010 - 2015." *Ministry of Education, Republic of Rwanda* , July 2010.

"Education Sector Strategic Plan (ESSP) 2013/14 - 2017/18." *Ministry of Education, Republic of Rwanda*, Oct. 2013.

Frick, Walter. "Here's Why People Trust Human Judgment Over Algorithms." *Harvard Business Review*, 27 Feb. 2015.

Gibbons, R. *Game Theory for Applied Economists*. Princeton University Press, 2013.

Glick, Peter, and David E. Sahn. "Early Academic Performance, Grade Repetition, and School Attainment in Senegal: A Panel Data Analysis." *The World Bank Economic Review*, vol. 24, no. 1, 2010, pp. 93–120.

Grant, Monica J., and Kelly K. Hallman. "Pregnancy-Related School Dropout and Prior School Performance in KwaZulu-Natal, South Africa." *Studies in Family Planning*, vol. 39, no. 4, 2008, pp. 369–382.

"How It Works." *EFL Global*, www.eflglobal.com/works/.

ICT in Education Policy. Ministry of Education, Republic of Rwanda, Apr. 2016, www.mineduc.gov.rw/fileadmin/user_upload/pdf_files/ICT_in_Education_Policy_approved.pdf.

Ikeda, Miyako, and Emma García. "Grade Repetition: A Comparative Study of Academic and Non-Academic Consequences." *OECD Journal: Economic Studies*, 2013.
"Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region." *UNICEF-ACER (Australian Council for Educational Research)*, 2016, www.unicef.org/esaro/UNICEF-ACER(2016)QualityofEducation-Rwanda-CS.pdf.
Jensen, Robert. "The (Perceived) Returns to Education and the Demand for Schooling*." *Quarterly Journal of Economics*, vol. 125, no. 2, 2010, pp. 515–548.
Jones, Sam. "How Does Classroom Composition Affect Learning Outcomes in Ugandan Primary Schools?" *International Journal of Educational Development Pages*, vol. 48, May 2016, pp. 66–78.
Kleinberg, Jon, et al. *A Guide to Solving Social Problems with Machine Learning*. Harvard Business Review, 8 Dec. 2016.
Kleinberg, Jon, et al. "Prediction Policy Problems." *American Economic Review*, vol. 105, no. 5, 2015, pp. 491–495.
Kuteesa, Hudson. *Smartphones to Be Used to Track Quality of Learning*. The New Times Rwanda, 15 Feb. 2017, www.newtimes.co.rw/section/article/2017-02-15/208007/.
Lee, Eun Whan. "Selecting the Best Prediction Model for Readmission." *Journal of Preventive Medicine and Public Health*, The Korean Society for Preventive Medicine, 31 July 2012.
Lincove, Jane Arnold. "The Influence of Price on School Enrollment under Uganda's Policy of Free Primary Education." *Economics of Education Review*, vol. 31, no. 5, 2012, pp. 799–811.
Manacorda, M. "The Cost of Grade Retention." *Federal Reserve Bank of St Louis*, 2010.
Marquez-Vera, C, et al. "Predicting School Failure Using Data Mining." 2013, educationaldatamining.org/EDM2011/wp-content/uploads/proc/edm2011_paper11_short_Marquez-Vera.pdf.
Montenegro, Claudio E., and Harry Anthony Patrinos. "Comparable Estimates of Returns to Schooling around the World." *World Bank Policy Research Working Papers*, Sept. 2014.
Morgan, Holly G. *Effects of Kindergarten Retention Policy on Children's Cognitive Growth in Reading and Mathematics*. Childhood Education, 2006.
Mullainathan, Sendhil, and Eldar Shafir. *Scarcity: the New Science of Having Less and How It Defines Our Lives*. New York, Picador/Henry Holt and Company, 2014.
Muralidharan, Karthik, et al. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *NBER Working Paper No. 22923*, 2016.
Nekatibeb, Techome. "Low Participation of Female Students in Primary Education: A Case Study of Dropouts from the Amhara and Oromia Regional States in Ethiopia." *UNESCO International Institute for Capacity Building in Africa*, May 2002.
Perry, Walter L, et al. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Safety and Justice Program, 2013.

"Positivo, Made in Rwanda Computer to Increase Country's Export." *Ministry of Trade and Industry, Republic of Rwanda*, 6 Nov. 2015, www.minicom.gov.rw/index.php?id=24&tx_ttnews%5Btt_news%5D=1010&cHash=edd69331bdbe03c21c0517365967bb46++.

Pritchett, L, and A Beatty. *The Negative Consequences of Overambitious Curricula in Developing Countries*. IDEAS Working Paper Series from RePEc, 2012.

Rosenthal, Robert, and Lenore F. Jacobson. "Teacher Expectations for the Disadvantaged." *Scientific American*, vol. 218, no. 4, 1968, pp. 19–23.

"Rwanda Vision 2020 (Revised 2012)." *Rwandan Development Board*, 2012.

Sara, Nicolae-Bogdan, et al. *High-School Dropout Prediction Using Machine Learning: A Danish Large-Scale Study*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2015.

Scher, Daniel. "The Promise of Imihigo: Decentralized Service Delivery in Rwanda, 2006-2010." *Innovations for Successful Societies*.

Schwerdt, Guido, and Martin West. "The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida." *HKS: Program on Education Policy and Governance Working Papers Series*, Aug. 2012.

"2015 Education Statistical Yearbook." *Ministry of Education, Republic of Rwanda*, June 2016.

Vuorikari, Riina, and Jonatan Castaño Muñoz, editors. *Research Evidence on the Use of Learning Analytics, Implications for Education Policy*. European Commission Joint Research Centre, 2016.

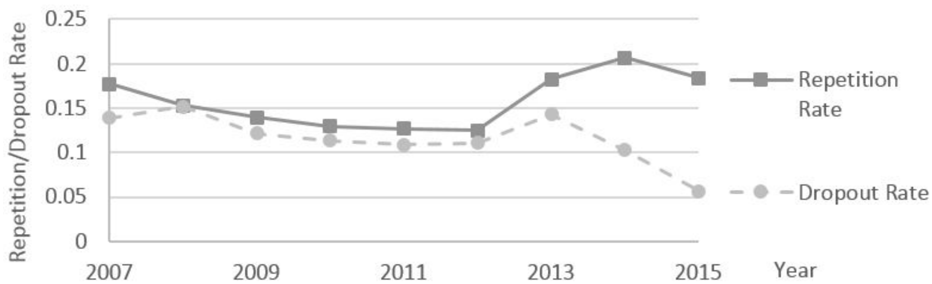
Williams, Timothy P., et al. "'Education at Our School Is Not Free' the Hidden Costs of Fee-Free Schooling in Rwanda." *Compare: A Journal of Comparative and International Education*, vol. 45, no. 6, 2015, pp. 931–952.

Woldehanna, Tassew, and Adiam Hagos. "Economic Shocks and Children's Dropout from Primary School: Implications for Education Policy in Ethiopia." *Africa Education Review*, vol. 12, no. 1, pp. 28–47., www.tandfonline.com/doi/ref/10.1080/18146627.2015.1036548.

10 Appendix

10.1 Figures

Figure A1: Repetition and Dropout from 2007 to 2015



10.2 More on Relevant Policies

10.2.1 Imihigo

The Minister of State issued instructions on "Determining Modalities for Performance Appraisal of Teachers"[73], in which the details of teacher Imihigo contracts are given. While teachers set their own targets (in line with the school targets), these instructions define 11 areas as a basis and criteria for teachers' appraisal:

1. Teacher preparation
2. Effective teaching methodologies and materials
3. Motivating students
4. Assessing and evaluating student progress
5. Coordination and supervision of students
6. Teamwork
7. Professionalism
8. Time management
9. Success rate of students
10. Cleanliness
11. Accountability

Head teachers have a slightly different set of criteria, including the following two:

- 12) "*ability to comply with the laws, orders and instructions governing education especially those relating to the promotion, repetition and dismissal of students*"
- 13) "*ability to fight anything that may lead students to dropout of school as well as how he/she try hard to bring back those who have left the school*"

[73] Instructions of Minister of State, *Determining Modalities for Performance Appraisal of Teachers*, N° 001/2016, 4/10/2016

Future versions of teachers' Imihigo contracts could also include criteria similar to these two, as suggested in our policy proposals.

10.3 More on Data and Variables

As mentioned in the main text, we use the Integrated Household Living Conditions Surveys[74] (EICV) for creating a prediction model and scorecard. EICV is conducted every 3 to 5 years and includes individual data on demographics, health, education, housing, services, economic data, etc. For individuals of school going age, data on both the current grade of enrollment as well as previous year's grade is recorded, allowing us to determine whether that person promoted, repeated, or dropped out. These are the outcome variables we predict using most of the EICV4 data. The following sections shortly go over data cleaning choices, variable categorization choices, and a list of all EICV variables we used in finding an optimal model.

10.3.1 Data Cleaning Choices

Going from raw EICV data to a dataset that (1) is in a format that can be used for training a prediction model and (2) will lead to sensible prediction logic, is a crucial first step. The following are a series of data cleaning decisions made to arrive at the final dataset:

- Being the focus of our study, we kept only individuals enrolled in primary, which left us with 13,453 observations. (In repetition prediction, we remove the dropout-observations, which leaves us with 12,576 observations: students who either promoted or repeated.)
- We replaced missing values with zero or sample average, whichever was more appropriate.
- We created some composite variables (e.g. *ontrack*, which is a function of *age* and *grade*) and collapsed several household member variables to the individual level (e.g. *hh_canread* was collapsed for family members as a measure of the pupil's household's literacy).
- Categorical variables cannot be directly used in a quantitative model, so a dummy was created for every value it takes. Of these new dummies, we removed the ones that represent a "missing" value, as we do not want to base scorecard prediction on any values "missing".

[74] The National Institute of Statistics Rwanda (NISR) is Rwanda's central agency for data collection and dissemination. The EICV4 data was retrieved from: http://microdata.statistics.gov.rw/index.php/catalog/75

- Several other variables are removed due to collinearity (which LASSO does not automatically handle)

After data cleaning, we end up with more than 300 variables from which to choose predictor variables for the eventual scorecard. The list of about 140 variables, with categorical variables not yet turned into dummies, is given further down.

Note also that when we use this data to train a repetition prediction model, we exclude observations of pupils who dropped out (and vice versa). The reason is that we do not want prediction of repetition to be (partly) based on the “prediction of not-dropping out”. This is also related to the paragraph *Variables to exclude from dropout predictions* in the next section

10.3.2 Variable Categorization

We categorize each variable along several dimensions. The list of variables in the next section includes 4 columns with such categorization.

Column 2 - Collection Level - The difficulty or cost tiers of variables are based on the lowest common denominator of the type of person who would be able to collect the data. These three groups of potential data collectors are:

1. <u>Teacher Variables</u>: These are variables that a teacher could reasonably collect from school records, observation, or asking the student in their class. Examples include: student age, grade, basic literacy, basic numeracy, etc
2. <u>Community Variables</u>: These are variables that a lay member of the community with basic training could collect by interviewing a student’s parent or guardian. Examples include: parent’s education level, parent’s expenditure on school items. Community variables also include all “Teacher Variables”
3. <u>Professional Variables</u>: These are variables that a skilled enumerator would be required to collect in a centralized manner (such as in a household interview). Examples include: variables that are combined into indexes, such as total household consumption. Professional Variables also include the “Teacher” and “Community” subsets.

The reason for this classification is the tradeoff between model accuracy and cost/feasibility of data collection and scorecard creation: more data, more accurate data, and more intricate data (i.e. the data collected at the “professional level”) generally leads to better prediction models.

The downside is the high cost. Teachers on the other hand are able to collect some basic variables in a fast, easy, and cheap way. The downside is possibly low prediction accuracy. The discussion and conclusions of this trade-off can be found in the main text.

Note that, after transforming categorical variables into dummies, we end up with 109, 284, and 325 predictor variables at respectively the teacher, community, and professional level.

Column 3 - Prediction Level - This categorization is important when considering *Absolute predictions* versus *Relative predictions*. Another choice in scorecard creation is whether 'absolute' prediction is important or whether 'relative' prediction is important. 'Absolute' risk predictions will include data both about the individual student as well as school and class level characteristics, such as whether a school is subsidized or not, urban or rural setting, etc. This will allow accurate comparisons of students' risk *across* schools. Relative predictions will only include data on student and family characteristics and can only be accurately used for *within* school comparison. Which option is preferable will depend heavily on the policy proposal and whether across school or within school differentiation is more important.

If a scorecard is meant to inform a teacher about which children *within* a class are relatively more at-risk, we do not need scores to be relatively comparable across schools or regions. In this case, regional or school level variables do not need to be included in the scorecard as they move the score of all pupils in the class or school up or down by the same amount.

One hybrid scorecard approach would be to create an administrative model that includes both individual level as well as school and regional level characteristics as predictors, but to create a scorecard that only includes the former. The predictors at the school and regional level[75], can be used centrally to calculate a "scorecard offset score" for each school in the country that could then be applied to all students from that school.

Column 4 - Variables to Exclude from Dropout Predictions - This column indicates the variables that are highly correlated with a dropout event *only after dropout occurred*. For example: any variable related school payments will be highly "predictable" of dropout in the EICV data, since pupils who dropped out of school no longer pay school fees. To try and

[75] Some of these are: "Is the school public or subsidized?", "Is the school in districts X, Y, or Z?", etc.

overcome these barriers when predicting dropout, we kept only those variables that plausibly do not change much when a student goes from "*in school*" to "*out of school*" status. However, even with removal of these variables, we still think the remaining EICV variables (similarly collected *after* the dropout event) are still not a good representation of the pupil's situation when he/she was still in school, *before* the dropout event. As such, the results of our dropout prediction analysis are not included in this document.

Column 5 - Categorical Variable - These are the EICV variables that take on different categories. For example, the variable *Relation with head of household* takes on several, non-ordered values. Before we can incorporate this variable into our model as a predictor, we create a new binary variable (i.e. a *dummy* variable) for each possible value this *Relation with head of household* variable can have. This implies that the pool of potential predictor variables we are looking at is much larger than the list below.

Figure A3: list of variables

Variable Name	Collection Level	Prediction Level	Exclude for dropout predict?	Catego-rical Variable	Variable Label
kigali	Teacher	School/class			kigali
p1	Teacher	School/class			in primary 1
p2	Teacher	School/class			in primary 2
p3	Teacher	School/class			in primary 3
p4	Teacher	School/class			in primary 4
p5	Teacher	School/class			in primary 5
p6	Teacher	School/class			in primary 6
urb	Teacher	School/class			urban area
district	Teacher	School/class		Yes	District
id1	Teacher	School/class		Yes	Province
s4aq10	Teacher	School/class	Yes		School has separate toilet facilities for boys and girls
schooltype	Teacher	School/class	Yes	Yes	public-1 or subsidized-3
age_y	Teacher	Individual			age
female	Teacher	Individual			female
hd_mobile	Teacher	Individual			head of hh hass mobile phone
hhsize	Teacher	Individual			household size
ontrack	Teacher	Individual			pupil is on track
prim_grade	Teacher	Individual			which grade: 1 to 6
s3q4	Teacher	Individual			Suffered from health problem in the last 4 weeks
s4bq3	Teacher	Individual			Can read a letter or a simple note
s4bq4	Teacher	Individual			Can write a letter or a simple note
s4bq5	Teacher	Individual			Can perform a written calculation
s4bq8	Teacher	Individual			Has a mobile telephone
upperprim	Teacher	Individual			Pupil is in upper primary dummy
s1q13	Teacher	Individual		Yes	Father still alive

s1q14	Teacher	Individual		Yes	Mother still alive
s1q2	Teacher	Individual		Yes	Relation with head of household
s3q2	Teacher	Individual		Yes	Suffers from a major disability
s4bq6	Teacher	Individual		Yes	Feels confident about using a computer
paidtuition	Teacher	Individual	Yes		paid any tuition
s4aq13	Teacher	Individual	Yes		Missed any day at school in the last 7 days
s4aq14	Teacher	Individual	Yes		Number of school days missed in the last y days
s6eq13	Teacher	Individual	Yes		Ever worked before
s6fq1	Teacher	Individual	Yes		Fetched water last 7 days
s6fq11	Teacher	Individual	Yes		Did other household chores last 7 days
s6fq3	Teacher	Individual	Yes		Foraged for firewood last 7 days
s6fq5	Teacher	Individual	Yes		searched fodder last 7 days
s6fq7	Teacher	Individual	Yes		Went to the market last 7 days
s6fq9	Teacher	Individual	Yes		Cooked for household last 7 days
hd_canread	Community	Individual			head of hh can read simple note
hd_disable	Community	Individual			head of hh is disabled
hd_ed_lopr	Community	Individual			head of hh has some lower primary education
hd_ed_lose	Community	Individual			head of hh has some lower secondary education
hd_ed_none	Community	Individual			head of hh has no education
hd_ed_popr	Community	Individual			head of hh has some post primary education
hd_ed_uni	Community	Individual			head of hh has some university education
hd_ed_uppr	Community	Individual			head of hh has some upper primary education
hd_ed_upse	Community	Individual			head of hh has some upper secondary education
hd_edulvl	Community	Individual			head of hh education level - 1 to 8 blocks
hd_female	Community	Individual			head of hh is female
hd_unemp_long	Community	Individual			head of hh is long term unemployed
hd_unemp_shrt	Community	Individual			head of hh is short term unemployed
s1q7	Community	Individual			Absent in the last 12 months
s2q2	Community	Individual			Always lived in the district
s2q3m	Community	Individual			Time since last move (months)
s2q3y	Community	Individual			Time since last move (years)
s2q6	Community	Individual			Duration of stay in the previous location (years)
s3q6	Community	Individual			Medical consultation over the last 4 weeks
ubudehe_1	Community	Individual			ubudehe category 1
ubudehe_12	Community	Individual			ubudehe category 1 or 2
ubudehe_123	Community	Individual			ubudehe category 1 or 2 or 3
s1q10	Community	Individual		Yes	Lived in another HH during absence
s1q12	Community	Individual		Yes	Reason for absence
s1q8	Community	Individual		Yes	Length of absence in last 12 months
s1q9	Community	Individual		Yes	Reason for absence
s2q5	Community	Individual		Yes	Kind of place lived prior to arrival
s2q7	Community	Individual		Yes	Primary reason for moving
s2q8	Community	Individual		Yes	Moved alone or with other HH members
s3q3	Community	Individual		Yes	Main health insurance
s3q5	Community	Individual		Yes	Nature of the health problem
s3q7	Community	Individual		Yes	Reason for medical consultation
paid_contrib	Community	Individual	Yes		paid any contributions
paid_material	Community	Individual	Yes		paid any contribution
paid_nontuition	Community	Individual	Yes		paid any non-tuition costs
paid_trnsbrd	Community	Individual	Yes		paid any board or transport

paid_uniform	Community	Individual	Yes		paid any uniform costs
s4bq7	Community	Individual	Yes		Attended literacy course
s6aq2	Community	Individual	Yes		Work on own/ HH farm (last 12 months)
s6aq3	Community	Individual	Yes		Agricultural Work for salary, wages, in-kind (last 12 months)
s6aq4	Community	Individual	Yes		Non-farm work for salary or wages (last 12 months)
s6aq5	Community	Individual	Yes		Non-farm business for cash/profit (last 12 months)
s6aq6	Community	Individual	Yes		Worked in HH member non-farm business for no pay
s6aq7a	Community	Individual	Yes		Number of jobs over past 12 months
s6aq7b	Community	Individual	Yes		Number of jobs over past 7 days
s6eq1	Community	Individual	Yes		Worked for one hour in the last 7 days in S6A activities
s6eq10	Community	Individual	Yes		Desire to work at the present
s6eq17	Community	Individual	Yes		Wanted additional work
s6eq2	Community	Individual	Yes		Usually works but absent last week
s6eq4	Community	Individual	Yes		Usual hours worked in job; 7 day period
s6eq7	Community	Individual	Yes		Able to start work if the opportunity existed
s6eq8	Community	Individual	Yes		Seeking a job during the last 4 weeks
s6fq10	Community	Individual	Yes		Number of hours cooking last 7 days
s6fq12	Community	Individual	Yes		Number of hours spent on other household chores last 7 days
s6fq2	Community	Individual	Yes		Number of hours fetching water last 7 days
s6fq4	Community	Individual	Yes		Number of hours foraging for wood last 7 days
s6fq6	Community	Individual	Yes		Number of hours searching fodder last 7 days
s6fq8	Community	Individual	Yes		Number of hours going to market last 7 days
s4aq12	Community	Individual	Yes	Yes	Who paid school expenses over the last 12 months
s4aq15	Community	Individual	Yes	Yes	Cause of not attending school
s4aq9	Community	Individual	Yes	Yes	Problem experiences at school
s6aq8	Community	Individual	Yes	Yes	Main reason for not working (last 12 months)
s6eq3	Community	Individual	Yes	Yes	Reason for absence from work: last 7 days
s6eq5	Community	Individual	Yes	Yes	Worked in other activities in the last 7 days
s6eq6	Community	Individual	Yes	Yes	Reason not workin in the last 7 days
ln_cons1	Professional	Individual			aggregate consumption (log)
ln_cons1_ae	Professional	Individual			aggregate consumption / ae (log)
ln_consumption	Professional	Individual			consumption (log)
ln_exp10	Professional	Individual			Provision of house (log)
ln_exp11	Professional	Individual			Other benefits (log)
ln_exp12	Professional	Individual			Annual non-food expenditures (log)
ln_exp12_wh	Professional	Individual			Annual non-food expenditures (without health) (log)
ln_exp13	Professional	Individual			Monthly non-food expenditures (log)
ln_exp14_1	Professional	Individual			Frequent non-food expenditures without filter question 0 (log)
ln_exp14_2	Professional	Individual			Frequent non-food expenditures without filter questions (log)
ln_exp15_1	Professional	Individual			Food expenditures without filter question 0 (log)
ln_exp15_2	Professional	Individual			Food expenditures without filter question (log)
ln_exp16_1	Professional	Individual			Own food consumption without filter question 0 (log)
ln_exp16_2	Professional	Individual			Own food consumption without filter question (log)
ln_exp16a_1	Professional	Individual			Own non-food consumption with filter question 0 (log)
ln_exp16a_2	Professional	Individual			Own non-food consumption without filter question (log)
ln_exp17	Professional	Individual			Use value of durable goods (log)
ln_exp18	Professional	Individual			Received transfers (log)
ln_exp4	Professional	Individual			Imputed rents (log)
ln_exp5	Professional	Individual			Actual rents (log)
ln_exp6	Professional	Individual			Maintenance costs (log)
ln_exp7	Professional	Individual			Water expenses (log)

ln_exp8	Professional	Individual			Electricity expenses (log)
ln_exp9	Professional	Individual			In-kind payments (log)
ln_sol_jan	Professional	Individual			aggregate consumption / ae Jan14 = 100 (log)
lncons_ae	Professional	Individual			ln aggregate consumption / ae Jan14 = 100
pov	Professional	Individual			Total Poverty
pov_ext	Professional	Individual			Extreme poverty
poverty	Professional	Individual			poverty category
quintile	Professional	Individual		Yes	income quintile
ln_costsboard	Professional	Individual	Yes		Costs for Boarding (log)
ln_costsboardtransp	Professional	Individual	Yes		Costs for Boarding or transport (log)
ln_costscontributions	Professional	Individual	Yes		Costs of contributions (log)
ln_costsmaterial	Professional	Individual	Yes		Costs of materials (log)
ln_costsnontuition	Professional	Individual	Yes		Costs of all non-tuition expenses (log)
ln_costsother	Professional	Individual	Yes		costs of other school expenses (log)
ln_coststotal	Professional	Individual	Yes		costs of total school expenditure (log)
ln_coststotal	Professional	Individual	Yes		Costs of schooling (log)
ln_coststransport	Professional	Individual	Yes		costs of transportation (log)
ln_coststutition	Professional	Individual	Yes		costs of tuition (log)
ln_costsuniform	Professional	Individual	Yes		costs of uniform (log)
ln_exp_edu_ae	Professional	Individual	Yes		log of per adult equiv. education expenses
ln_exp1	Professional	Individual	Yes		Education expenses (log)
totalcosts	Professional	Individual	Yes		costs of all school expenses

10.4 More on Quantitative Techniques

10.4.1 Machine Learning

Unlike normal regression techniques where we use the whole dataset to learn about how variables are related, machine learning approaches typically split analysis into two phases, "training" and "testing". Before any analysis is conducted, a dataset is split into two groups for these two different phases. In the training phase, a prediction model is built based on the training portion of the data set. Next, the results of this model (what variables are included, how they should be weighted, etc.) are applied to the testing data to make a prediction - what value is the outcome variable (i.e. dropout or repetition in our case) most likely to be.

The model is then evaluated by how well the predicted outcome matches up with the true observed outcome. This is called "out-of-sample" model performance and is a measure of how well an algorithm applies to "new" data. This is an important idea in a public policy context: it is important to ensure that the predictions we make are not only accurate for the data we initially analyze, but also for children that the analysis is later applied to.

10.4.2 Model choice: Logistic Regression

For the prediction analysis presented in this report, we use a logistic regression, also called a logit model. Logit models are used when the dependent variable - i.e. the outcome variable of interest - is a categorical variable. The most common use of a logit model is when the outcome is binary: "yes" or "no", "true" or "false", "1" or "0", etc. In our case "1" represents a child repeating a grade (or dropping out, if we want to predict dropout), and "0" represents no repetition. The logit model allows us to estimate the probability that the outcome variable Y_i for an individual pupil i is "1" (i.e. "repetition"), given (or "conditional on") K covariates X_{ij}:

- The 'j' stands for covariate j, which goes from 1 to K. In our case the covariates X_{ij} are the "predictor variables" we will use to predict the repetition or dropout outcome. Figure A3 in Appendix 10.3.3 provides a full list of variables we take into consideration as predictor variables X_{ij}. Some examples are "age", "household head education", "poverty status", "grade", etc.
- The 'i' stands for a specific individual i, which in a sample size of N pupils goes from 1 to N.

This probability that outcome Y_i=1 is also equal to the expectation (or average) of the outcomes Y_i, given K covariates $\mathbf{X}_i = [X_{i1}, X_{i2}, X_{i3}, \ldots, X_{iK}]$. As a mathematical model, this is:

$$\mathbf{E[Y_i|X_i] = Pr(Y_i{=}1|X_i) = \Lambda(\beta_0{+}\beta_1 X_{i1}{+}\beta_2 X_{i2}{+}...{+}\beta_N X_{iK})}, \quad \text{with Logit: } \mathbf{\Lambda(t) = e^t / (1 + e^t)}$$

When we create the logit model, using the EICV data and statistical software packages in Python, we are calculating estimates for the model parameters β_j, i.e. $\beta_0, \beta_1, \beta_2, \ldots, \beta_N$. With these numeric values for all of the β_j's, we then use the above logit model to predict the probability of a *new* pupil *n* with *new* values for its predictor variables $X_{n1}, X_{n2}, ..., X_{nK}$: we plug in these covariate values into the mathematical model above, and calculate that pupil's probability of repetition. As is described in the next section, logistic regression is easily combined with LASSO for variable selection.

Note that we did consider alternative models, for example Random Forests, but:
(1) these are less easily interpretable and not easily turned into a straightforward scorecard, and

(2) these models did not perform as well as the logistics regression model. This finding is in line with other research predicting dropout[76].

10.4.3 LASSO for Variable Selection and Preventing Overfitting

One problem of having a large number of predictor variables arises when fitting a logistic (or any other) model to the training data: the model is *overfitted* and prediction performance on test data - i.e. "new" out-of-sample data - will not be as good. One way around this is to create a model with more conservative estimates of the model coefficients β_j as well as a model with *fewer* predictor variables and hence *fewer* β parameters. Another reason for going with a lower number of model variables is to (1) limit the required data collection efforts, and (2) have a simple model to create a simple scorecard. That is, we want to end up with a model with a limited number of variables.

This is addressed by using LASSO, or *Least Absolute Shrinkage and Selection Operator*. LASSO imposes an additional restriction on the model estimates β_j:

$$\sum_j |\boldsymbol{\beta}_j| = |\boldsymbol{\beta}_1|+|\boldsymbol{\beta}_2|+\ldots+|\boldsymbol{\beta}_K| < \mathbf{C}$$

C is an "allowance" parameter we set at a chosen constant value: setting it high will lead to many non-zero coefficients, higher β_j estimates, and hence a model close to a non-LASSO model. Setting C low means we won't "allow" for a lot of non-zero β_j coefficients: only a few β_j coefficients will be non-zero, the coefficients will be lower and be more conservative estimates, and the model will have fewer parameters. Setting C to 0 means all β_j's are 0 and the model predicts the mean value for all observations.

At any value of C, this coefficient "allowance" C will be used as efficiently as possible. By changing C we get *more complex* or *less complex* models (in the sense of *number of predictors*). Slowly increasing C to get logistic LASSO regression models with incrementally more nonzero coefficients is how Figure 3 in the main text is created.

[76] Aulck, L, et al. (2016)

10.5 More on Scorecard Creation and Performance

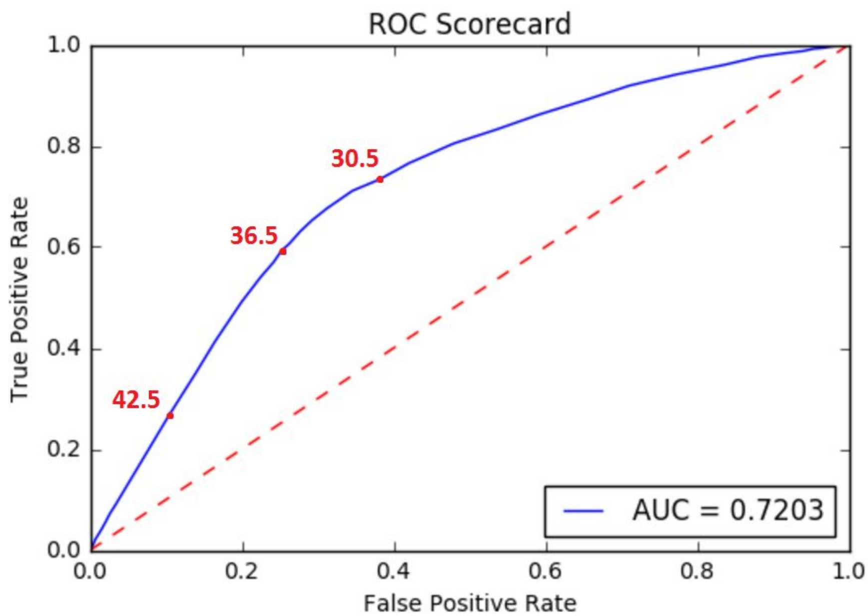
10.5.1 Model performance: AUC

The main text went into more depth on model performance as two measures to be minimized:

1. *Leakage*: Rate of non-repeating pupils misclassified as "at-risk"; i.e. *False Positive Rate*
2. *Undercoverage*: Rate of repeating pupils ***not*** identified as being "at-risk".

Minimizing *undercoverage* is equivalent to maximizing *coverage*, i.e. the *True Positive Rate.* This tradeoff between *False Positive Rate* (where "low is good") and *True Positive rate* (where "high is good") can be captured in one number, the AUC or *Area Under the Curve*. The curve-of-interest here is the ROC curve, or *Receiver Operating Characteristic* curve. The ROC curve is constructed by plotting the *False Positive Rate* versus *True Positive Rate* for every possible cutoff (or "threshold") used to classify predicted outcomes as either "1" or "0", i.e. "repeat" of "not repeat" in our case. A predicted outcome above the threshold is classified as "1". For a raw logit model the predicted outcome is a probability between 0 and 1, so the ROC is constructed by going over threshold range of 0 to 1. For our proof-of-concept scorecard, the predicted outcome variable lies between 0 and 64, so constructing the Scorecard ROC curve means changing threshold from 0 to 64 and plotting the *False Positive Rate* versus the *True Positive Rate* for each threshold value. This ROC curve is given in Figure (A3) below.

Figure A4: The ROC curve illustrating AUC and the tradeoffs made setting a threshold



Note the following about the above figure:

- The ideal prediction (i.e. the *perfect* model as in the upper left panel of Figure 9) would put us in the upper left corner of the *False-vs-True Positive Rate* space: The *True Positive Rate* is 1 (we correctly identified all repeaters correctly as being "at-risk"), and the *False Positive Rate* is 0 (we correctly identify none of the non-repeaters as being "at-risk").
- However, we are limited to the blue ROC curve: we can change the threshold and move *along* the ROC curve, but we cannot *go beyond* it. Three example threshold positions are shown on the above ROC curve; the same 3 thresholds used in the main text.
- Since we cannot *go beyond* the curve, we want the TOC curve to come as close as possible to the ideal point; the upper left corner where the *perfect* model would be. AUC is the number used to measure how close we get to that point: the greater the *Area Under the Curve*, the better our prediction model performs. Since the *False-vs-True Positive Rate* space is a 1-by-1 space, the maximum AUC equals 1 (i.e. the impossible *perfect* model). The red dotted line, with half of the area under the curve, is baseline performance: it is a random classifier, were we randomly classify anywhere between 0% and 100% of the pupils as being "at-risk". Hence any prediction model should have an AUC over 0.5. The closer to 1, the better.

- On the other hand, the bottom left corner (0, 0) and the upper right corner (1,1) are points on all possible ROC curves. This is because it is always possible to set the threshold so high so that no-one is predicted to be at risk (0% false positives, 0% true positives → lower left corner). At the other extreme, the threshold can be set so low that everyone is predicted to be at risk (100% false positives, 100% true positives → upper right corner).

Note also that;

- The AUC as model performance indicator was also used in figure 3.
- For a raw logit regression model, the thresholds would be values between 0 and 1.

10.5.2 Scorecard Creation Process

Running the logistic regression with our final variables and parameters gives us the regression table depicted below. With values for both the model coefficients β and the predictors X we can calculate ($\beta_0+\beta_1 X_{i1}+...+\beta_N X_{iK}$), which can take on any value. Applying Logit $\mathbf{\Lambda}$ to this value gives us a predicted probability between 0 and 1 for every observation. However:

(1) Applying a Logit function is not easily done on a (manual) scorecard and is less transparent.

(2) We are interested in a relative risk measure, not so much an absolute repetition probability.

(3) The values ($\beta_0+\beta_1 X_{i1}+...+\beta_N X_{iK}$) map one-to-one to the values between 0 and 1 that $\mathbf{\Lambda}(\beta_0+\beta_1 X_{i1}+...+\beta_N X_{iK})$ would give. So the ROC curve would be the same for both predictions.

Therefore we use only the linear part ($\beta_0+\beta_1 X_{i1}+...+\beta_N X_{iK}$) of the non-linear logit model $\mathbf{\Lambda}(...)$.

Figure A5: Raw Logit Regression Output

	coef	std err	z	P>|z|	[95.0% Conf.	Int.]
const	0.9771	0.534	1.831	0.067	-0.069	2.023
p1	-0.6145	0.159	-3.877	0.000	-0.925	-0.304
p2	-0.5279	0.148	-3.571	0.000	-0.818	-0.238
p3	-0.0513	0.140	-0.366	0.714	-0.326	0.223
p4	0.2475	0.138	1.794	0.073	-0.023	0.518
p5	0.6663	0.136	4.912	0.000	0.400	0.932
age_y	-0.0842	0.013	-6.566	0.000	-0.109	-0.059
s6fq1_1	0.5719	0.378	1.511	0.131	-0.170	1.314
s4aq15_2	0.3650	0.092	3.956	0.000	0.184	0.546
s4aq15_10	0.8204	0.176	4.665	0.000	0.476	1.165
s4bq8_1	-0.7198	0.222	-3.240	0.001	-1.155	-0.284
s6fq1_0	0.2492	0.383	0.651	0.515	-0.501	0.999
s6eq1_1	-0.1290	0.087	-1.480	0.139	-0.300	0.042

```
s4aq15_5         0.5631      0.096      5.882      0.000          0.375     0.751
s6aq6_0         -0.5455      0.309     -1.767      0.077         -1.151     0.060
s4bq3_1         -0.7674      0.130     -5.900      0.000         -1.022    -0.512
s4bq4_1         -0.7142      0.130     -5.473      0.000         -0.970    -0.458
s4bq5_1         -0.3299      0.084     -3.942      0.000         -0.494    -0.166
district_33     -0.6173      0.165     -3.741      0.000         -0.941    -0.294
district_34      0.8345      0.108      7.713      0.000          0.622     1.047
district_35      0.5607      0.115      4.875      0.000          0.335     0.786
district_41     -0.4901      0.150     -3.271      0.001         -0.784    -0.196
id1_2            0.2019      0.054      3.736      0.000          0.096     0.308
id1_4           -0.2139      0.068     -3.135      0.002         -0.348    -0.080
schooltype_3     0.1384      0.047      2.927      0.003          0.046     0.231
paidtuition     -0.6171      0.122     -5.073      0.000         -0.856    -0.379
==================================================================================
```

Here, we shortly go over the steps to turn this model ($\beta_0+\beta_1 X_{i1}+...+\beta_N X_{iK}$) into a scorecard:

- For positive coefficients on "yes/no" type variables, round the coefficient to the nearest integer. On the scorecard, add that integer value to the total score if the answer is "yes".
- For negative coefficients on "yes/no" type variables, take the absolute value of the coefficient and round it to the nearest integer. On the scorecard, add that integer value to the total score if the answer to the question is "no". The idea is to only have to add, not subtract numbers.
- For age and grade, which each take on multiple values, create matrix and add coefficients for each age-grade combination. Take the absolute value of the most negative value and add it to all other values in this table. Again, this results in only positive numbers to add to total score.
- Combine multiple dummies of the same categorical variable into a single question.

10.5.3 Potential Changes to Model and/or Scorecard

Our analysis and scorecard creation can possibly be improved upon, both in terms of performance and in terms of transparency and acceptance:

- While we explored interaction and higher order terms, model performance did not improve. This is probably because higher order features add more potential for overfitting during model creation (worsening out-of-sample prediction) while adding little additional *information* to the data. However, a deeper analysis would be required.
- If repetition dynamics are very different for different groups, e.g. ages, grades, etc., then fitting different models to these subgroups could be explored to improve performance.

- Some scorecard questions might be a bit sensitive (e.g. “Has the child worked...”). Additional analysis could uncover substitute question explaining a similar amount of variation. Moreover, some questions - while providing accurate prediction - might not be intuitively related to repetition (or dropout), which could undermine credibility of the scorecard. Further analysis could try and replace these questions.

10.6 More on Implementation - Stakeholder Mapping

In considering implementation options, it is important to understand the dynamics and relations between stakeholders at all levels in the education system. Figure A7 below gives a visual mapping while Figure A6 gives a short overview of each stakeholder’s responsibilities and actions, as well as effect.

Figure A6: Stakeholder Mapping and Key Responsibilities

Actor	**Responsibilities and Actions**	**Effects on Repetition and Dropout Outcomes**
Ministry of Education (MINEDUC)	Planning and Policies Standards Collecting data, and monitoring and evaluation	Through high level policy making, instructions, and directives to implementing actor, MINEDUC’s position and actions highly affect outcomes.
Ministry of Public Service and Labour (MIFOTRA)	Public service staffing Pay and wages	Indirect impact on repetition and dropout through teacher quality and teacher incentives.
Ministry of Local Government (MINALOC)	Decentralization of service delivery and Setting “imihigo” (performance contracts)	Indirect effect through structure and hierarchy in educational service delivery. Also sets incentives of local actors (e.g. Mayors) through “imihigo” details.
Rwanda Education Board (REB)	Implementing agency of the MINEDUC; manages school system operations	Curriculum development: impacts difficulty and hence repetition and dropout. Teacher Development: impacts teach quality, norms, and culture, which affect repetition and dropout Central examination: sets academic expectations towards students and school, affecting dynamics of repetition and dropout.
REB’s Regional Inspectors (RIs)	Interact directly with School Officials, and officials from Local Governments.	Inspection and assessment of schools, including implementation of educational legislation, policies, programs, and other instructions from the Ministry: the quality of this affects how well policies and instructions improve school performance, including on indicators like repetition and dropout.
District Education officers (DEOs)	planning and budgeting, operational plans, monitoring school management, recruitment and deployment of	As the main decentralized unit of managing functions of day-to-day delivery of education services, this “institution” is crucial in linking central policymaking (MINEDUC) and central operationalizing (REB) with actual education service delivery at the District and

	teachers, providing education statistics	Sector Level. Through the actors below, there is a huge potential impact on outcomes like repetition and dropout.
District Mayor	Education only part of responsibilities (e.g. directly Hires School Headmasters)	Through "Imihigo" has incentive to get schools to perform well, e.g. lower repetition and dropout. Also oversee DEOs to operationalize education in district
Sector Education Officers (SEOs)	collect and aggregate education data and support supervision of schools	As a crucial practical link between policies and instructions from above and the actual schools in the sector, this actor's performance can highly influence the sector's schools' performance.
Parent Teacher Associations (PTAs)	Among others: decide on bonuses for teachers locally	Impact on teacher incentives and quality, leading to impact on repetition and dropout.
Schools and teachers	The ultimate responsibility of delivering educational services	All the above impacts teachers' quality, incentives, norms, culture, etc. – as well as student expectations – impacting school performance on e.g. repetition and dropout.

*Figure A7: Stakeholder Map Visualizatio*n