Baseline Report

Evaluation of the Secondary Education Activity of the Morocco Education & Training for Employability Project

April 15, 2020
Revised February 16, 2021
Revised September 20, 2021
Revised September 22, 2022

Catalina Torrente, Paolo Abarcar, Audrey-Marie Moore, Margo Berends, Galina Lapadatova, Dara Bernstein, Emilie Bagby and Matt Sloan

Submitted to:

Millennium Challenge Corporation
1099 Fourteenth St, NW
Suite 700
Washington, DC 20005
Project Officers: Carolyn Perrin and Ryan Moore
Contract Number: 95332418A0018

Submitted by:

Mathematica
1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

This page has been left blank for double-sided copying.

Contents

Tables

Figures

Acronyms

AAER	Association for Supporting School Success *(Association d'appui à l'école de la réussite)*
AREF	Regional Academy of Education and Training *(Académies Régionale de l'Education et la Formation)*
ASS	Sports Association Budget *(Association Sportive Scolaire)*
BFI	Big Five Inventory
DEUG	First university cycle (*Diplôme d'études universitaires générales*)
DI	Data Ingénierie
ECA	Extracurricular activities
EMIS	Education Management Information Systems
FM	Fès-Meknès
GoM	Government of Morocco
ICC	Intra-class correlation
IQ	Intelligence quotient
IRB	Institutional review board
ISIM	Integrated School Improvement Model *(Attahadi Model)*
LS	Lower secondary
MAD	Moroccan dirhams
MASSAR	Assessment of Student Achievement and Information System (*L'évaluation des Acquis des Élèves et du Système d'Information)*
MCA-M	Millennium Challenge Account Morocco
MCC	Millennium Challenge Corporation
MDE	Minimum detectable effect
MENFPESRS	Ministry of National Education, Vocational Training, Higher Eduction and Scientific Research (*Ministère de l'Education Nationale, de la Formation Professionnelle, de l'Enseignement Supérieur et de la Recherche Scientifique)*
MS	Marrakech-Safi
OECD	Organisation for Economic Co-operation and Development
O&M	Operations and Maintenance
PISA	Program for International Student Assessment

PNEA	National Student Assessment Program *(Programme National d'Evaluation des Acquis des Elèves)*
RCT	Randomized controlled trial
SBM	School-based management
SIP	School improvement project (*Projet d'établissement intégré [PEI]*)
STEM	Science, technology, engineering, and mathematics
TIMSS	Trends in International Mathematics and Science Study
TTH	Tanger-Tétouan-Al Hoceima
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	Upper secondary
USD	United States dollars
USAID	United States Agency for International Development
WASH	Water, sanitation, and hygiene

I. Introduction

Youth unemployment is currently a global policy issue. More than 71 million youth (ages 15–24) are unemployed and more than 142 million are out of school (UN 2018). Nearly one billion youth are expected to enter the workforce in the next decade and will likely face irregular work or work in the informal sector (UN 2018). Unemployment and underemployment often lead to migration, extreme poverty, and, in some cases, engagement in extremism (UN 2018). Individual unemployment at an early age can also have a negative effect on future earnings and increase the likelihood of joblessness later in life (S4YE 2015).

In Morocco, youth ages 15 to 29[1] account for 80 percent of the country's unemployed (USAID 2019). Youth unemployment in Morocco has been on the rise in the past decade and was estimated at 26.7 percent in the third quarter of 2019, two percentage points below the highest value recorded since 2010 (Trading Economics 2019). Low quality of secondary education, low rates of participation and completion at the secondary and tertiary levels, and misalignment between youth skills and the needs of the private sector are major factors negatively impacting the employability of Moroccan youth (MCC 2015b). Although net enrollment in primary education was nearly 100 percent in 2018, it was 65 percent in secondary education and 36 percent in tertiary education (UNESCO 2019). Further, youth who attend school are not acquiring the basic skills they need to do well academically and to participate in the labor market. Morocco ranked 75th among 79 countries on the 2018 Program for International Student Assessment (OECD 2019) and in the bottom four among 39 participating countries in the 8th grade math and science 2015 Trends in International Mathematics and Science Study (Mullis et al. 2016a, 2016b).

In 2015, the Government of Morocco (GoM) and MCC signed a second compact in the amount of $450 million aimed at addressing the country's major binding constraints to economic growth, one of which is low education quality. The Education and Training for Employability Project ($217 million) is one of two major investments under the compact. It encompasses a Secondary Education Activity, which is the focus of this report, and a Workforce Development Activity. [2] The Secondary Education Activity aims to test a new model for public secondary schools. The activity consists of three subactivities: (1) Integrated School Improvement Model (ISIM or *Attahadi* model in its Moroccan name) (which "equips youth with the skills the modern labor market demands, in a way that is cost-effective, sustainable, and scalable" [MCC 2015a]), (2) Student Assessment and Education Management Information System, and (3) School Infrastructure and Equipment Operations and Maintenance.

[1] Please note that donors working in international development use different age categories to define youth. USAID (referenced here) considers the youth population to be ages 15-29 while the UN (referenced in the previous paragraph) considers youth, people ages 15-24. Several donors consider youth to be ages 10-29.

[2] The Workforce Development Activity aims to improve the quality and relevance of, and equitable access to, private sector-driven technical and vocational education and training (MCC 2015a). The compact is also investing in a Land Productivity Project ($168 million) aimed at improving governance of the land sector and efficiency and inclusivity in the ownership of rural land, and optimizing the development and management of industrial land. The Workforce Development Activity and Land Productivity Project are the subject of separate evaluations.

MCC contracted Mathematica to conduct an independent evaluation of the Secondary Education Activity. The evaluation design uses a mixed-methods approach to assess the effects of the activity and includes (1) a randomized controlled trial (RCT) of the ISIM/Attahadi subactivity (including a cost analysis) and (2) an implementation study of all three subactivities.

This report presents baseline findings for the RCT of the ISIM/Attahadi subactivity. Analyzing baseline data for the RCT serves two purposes: (1) to describe the schools, teachers, and students eligible to participate in the activity, and (2) to assess whether randomization produced equivalent study groups. Although random assignment ensures that treatment status is not a result of specific characteristics, it could still result in chance differences between intervention groups on characteristics that might be correlated with the outcomes that the intervention is seeking to affect. Using baseline data allows us to check for those chance differences and, should they arise, to adjust for them in the final analysis. We provide context for the RCT and describe the baseline findings in the chapters that follow.

II. The Secondary Education Activity

The overarching goal of the Secondary Education Activity is to equip secondary school students with skills that are relevant to the private sector, create a more employable workforce, and galvanize economic growth. We describe the subactivities, beneficiaries, and geographic scope of the Secondary Education Activity in this section. The activity comprises three subactivities, summarized in Table II.1.

Table II.1. Subactivities in the Secondary Education Activity

Subactivity	Interventions	Geographic scope
1. Integrated School Improvement Model (ISIM / *Attahadi model* [a])	• Training programs to improve school leadership and to support regional level education staff to implement the reform package • Development and use of school improvement projects to improve quality of school management and decrease inequity at the school level • Infrastructure improvements (for example, adding classrooms and water, sanitation and hygiene facilities) • Delivery of equipment to schools including (1) IT equipment for administrative offices and computers for multipurpose rooms (2) didactic equipment, and (3) additional equipment for school clubs and what was requested under each school's school improvement project • Training programs to improve teachers' pedagogical practices. This includes modules on the use of information and communications technology, soft skills, management and leadership, assessment and accountability, prevention of violence, extracurricular activities, among other things. Training is expected to be rolled out over a full school year in program schools.[b] • The Education for Employability Partnership Fund for NGOs to implement various additional activities in schools based on partner's proposals	90[c] schools across Tanger-Tétouan-Al Hoceima (TTH), Fès-Meknès (FM), and Marrakech-Safi (MS) The subactivity started in the TTH region and then expanded to the other two regions
2. Student assessment and education management information system (EMIS/MASSAR)	• Building GoM's capacity to analyze results of international and national student assessments and improving Morocco's National Student Assessment Program (PNEA) • Support for the use of data to inform GoM decision making • Technical assistance to improve the Moroccan Education Management Information System (EMIS), known as MASSAR, for the use of decision-makers at the local, regional, and national levels • Improved continuous monitoring and remediation practices. • Improved initial teacher training in assessment. • Development of a school performance measurement framework.	National

Subactivity	Interventions	Geographic scope
3. The school infrastructure and equipment operations and maintenance (O&M)	• Technical assistance to develop a new approach to school infrastructure development, operations, and maintenance • Capacity building for regional actors engaged in the Attahadi subactivity to support implementation of the new approach • Pilot testing of the use of performance contracts to maintain and operate school infrastructure and information technology	TTH[d]

[a]This subactivity was formerly known as the Integrated School Improvement Model (ISIM).

[b]We are unclear about the length of each module and the details of continued support to be provided to teachers after training is completed.

[c]This number includes 6 schools that received the pilot program in TTH; 84 schools will be a part of the evaluation.

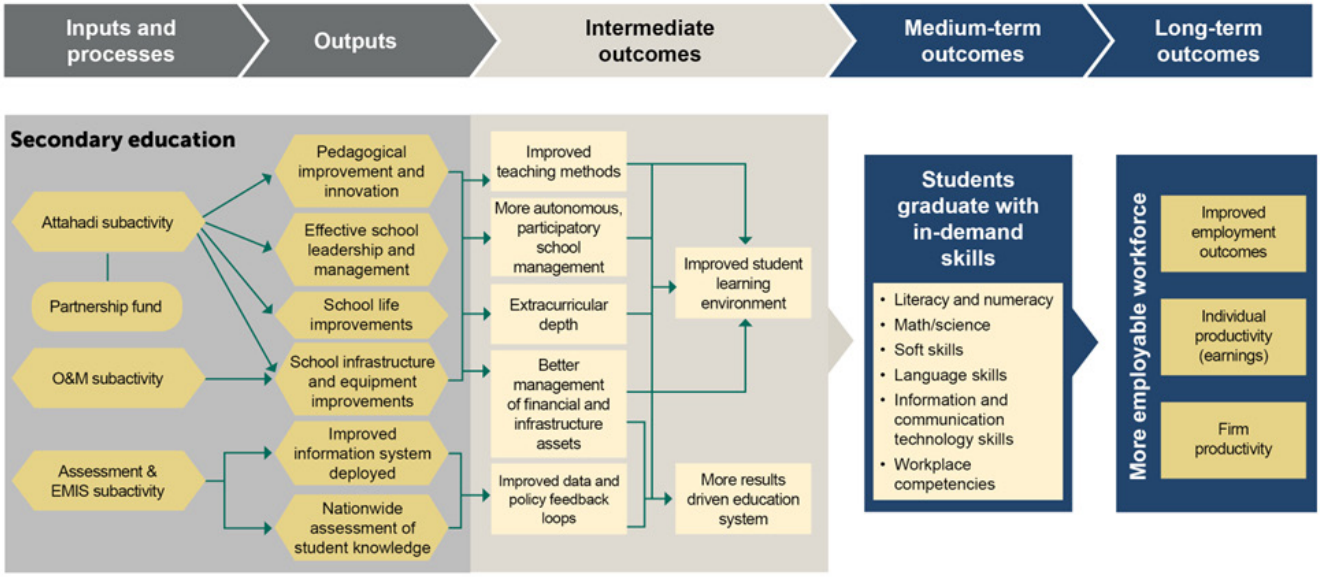
[d]The O&M subactivity was envisioned to be national in scope but has been scaled back to be delivered only to TTH Attahadi mode schools by the end of the compact because schools in TTH benefited from the rehabilitation and equipment before schools in the two other regions.

A. Program Logic

The program logic for the Secondary Education Activity encompasses a series of hypothesized causal links among program inputs and outputs and short-, medium-, and long-term outcomes that are expected to support the overarching goal of improving youth's preparation for entry into the workforce (Figure II.1). Each of the links in the theory of change represents an assumption by the activity's designers about how the interventions will affect the students in treatment schools and their families.

The program logic assumes that comprehensive educational reform will improve teacher performance in the classroom and create a more autonomous and participatory school management system through school improvement projects, capacity building, pedagogical innovations, data-driven decision making, and infrastructure and school life improvements. The changes will lead to a more results-driven education system and an improved student environment, which, in turn, are expected to increase student retention and learning. In the medium term, students will gain stronger foundational knowledge in literacy, math, science, and soft skills (including critical thinking, problem solving, and improved decision making). Once the students acquire these important practical skills, they will be better positioned to meet the needs of the private sector. The long-term outcomes of the activity include increased employability, workforce productivity, and earnings. The project activities will work directly with regional and provincial education actors, school directors, and teachers, who directly affect outcomes at the school, teacher, and student levels. Given the long-term nature of measuring improved employment outcomes and productivity, Mathematica will measure student completion and learning as the relevant interim outcomes to improved employment outcomes in this evaluation.

Figure II.1. Program logic for the Morocco Education and Training for Employability Project, Secondary Education Activity



Source: Millennium Challenge Account Morocco (MCA-M), modified by Mathematica.

Note: To take into account changes in project implementation, MCA-M is currently developing a revised logic framework. Mathematica's subsequent work on the evaluation will take into account the revised framework.

B. Program beneficiaries and geographic coverage

The primary beneficiaries of the Secondary Education Activity are the graduates of intervention schools and their families over a 20-year period commencing when the compact entered into force. The number of secondary students expected to graduate in these schools over the 20-year period is 376,600, based on estimates made by MCC in 2015. The MCC compact documents assume that the average family size in Morocco is 4.63, for a total number of approximately 1.7 million beneficiaries (MCC 2015a).[3] Even though primary beneficiaries are students in intervention schools and their families, teachers, school directors, the school system, and communities at large will also likely benefit from the Secondary Education Activity. These actors, however, are not considered program beneficiaries per MCC's guidelines for Economic and Beneficiary Analysis.

The Attahadi subactivity will be implemented in 90 lower secondary and upper secondary schools (including six pilot schools) across the three target regions.

C. Implementation and evaluation timeline

The Attahadi model implementation schedule and the timeline for the RCT are shown in Figure II.2. Baseline survey data for the RCT were collected in May 2018 for schools, teachers, students

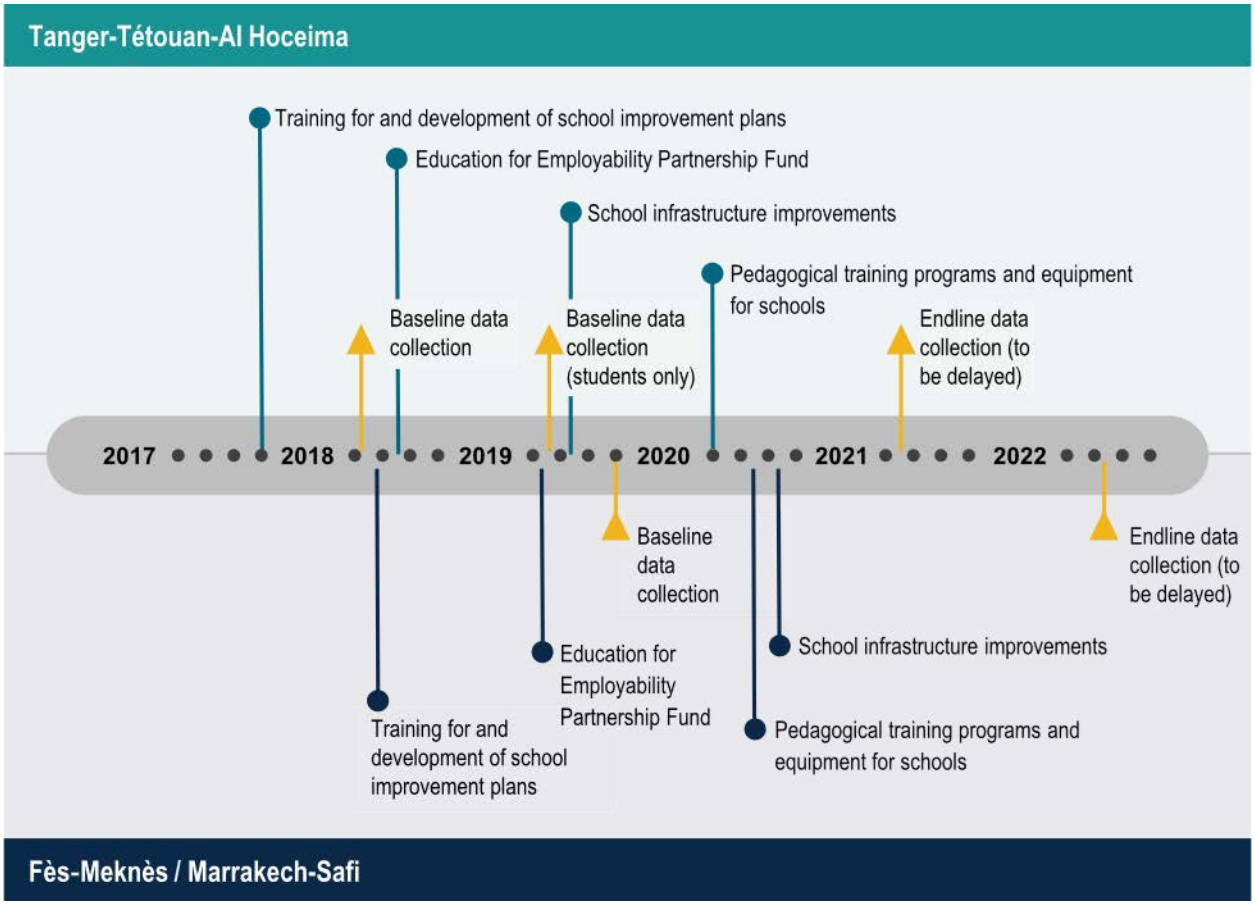
[3] MCC is in the process of updating these numbers with more recent data.

in TTH and again in May 2019 for a new cohort of students, and in November 2019 in FM/MS.[4] Baseline data were collected after some Attahadi model interventions had begun (namely the school improvement process and the Education for Employability Partnership Fund), but before infrastructure work began and teachers had received pedagogical training. Although baseline data collection occurred after the school improvement process and the Education for Employability Partnership Fund had begun, we did not expect the two activities to have been active long enough to affect student outcomes, including enrollment, and teacher outcomes. However, they might have started to affect some school-level outcomes at the time of data collection, which we explore in this report.

The plan was to collect endline data in April/May 2021 and 2022 in TTH and FM/MS, respectively, but we have decided to postpone endline data collection because of implementation delays arising from the COVID-19 pandemic and other factors. We are working with MCC to determine the appropriate timing of follow-up data collection activities. We will reflect any changes to our proposed evaluation design in a future updated Evaluation Design Report, as appropriate.

[4] Baseline data collection occurred twice in TTH because we decided to follow the next cohort of students for the evaluation after learning about implementation delays. Section IV.B provides further detail about this decision. Baseline data for schools and teachers however was only collected in May 2018.

Figure II.2. Attahadi subactivity and RCT timeline



This page has been left blank for double-sided copying.

III. Literature Review for the Attahadi Subactivity

This section summarizes the literature related to the main outputs of the Attahadi subactivity (pedagogical improvement and innovation, effective school leadership and management, school infrastructure and equipment improvements, and school life improvements), the outputs' association with intermediate and medium-term outcomes, and the potential contributions of this evaluation.

A. School-based leadership and management and student outcomes

Rigorous evidence is limited on the links between school-based management (SBM) and improvements in student learning. SBM reforms vary in the level of autonomy given to the school and the degree of involvement by other stakeholders (such as local authorities, parents, and other members of the community), making it difficult to reach broad conclusions. However, most high-performing countries on international standardized assessments give local authorities and schools substantial autonomy over what is taught or over the allocation and management of resources (Bruns et al. 2011). In addition, evidence from low- and middle-income countries (LMICs) suggests that SBM can have positive impacts on school quality, access, and test scores. In Kenya, a randomized trial showed that the Extra Teachers Program, which used school-based management to work through parent-teacher associations to hire and train local contract teachers, improved student test scores, decreased teacher absenteeism, and led to a small change in student dropout (Duflo et al. 2007). The school committees were responsible for hiring the teachers (who had similar qualifications to the civil service teachers) and decided on contract extensions based on performance. The reductions in class size and the role of the school committee in monitoring and hiring teachers contributed to the positive impacts of the program. In Mexico, the AGE 125 program led to increased participation and reduced dropout in the first year and improved reading scores in the second year (Gertler et al. 2010). The AGE 125 program involved parents directly in school management by financing the parent associations and involving parents directly in the management of school grants.

The evidence suggests that it can take time to effect change in the functioning of schools and student learning, though the required duration and level of exposure to affect change is still an under-researched area. An SBM program in the Philippines that included funding for school improvement projects increased the average national achievement test scores over the course of three years (Yamauchi 2014). Based on experience from the U.S., a World Bank review of the literature suggests that SBM reforms need about five years before any fundamental changes occur at the school level; only after eight years of implementation can changes be seen in indicators such as student test scores (Bruns et al. 2011). It is possible to see changes in teacher practices and uptake in 1-2 years, but the uptake depends on the frequency, intensity, and duration of training and support.

B. Improvements to school infrastructure

Several types of school infrastructure interventions are known to contribute to improvements in student enrollment, attendance and learning. Apart from the construction of new schools, these interventions include the construction of additional spaces to expand and improve already existing schools and refurbishing of existing school buildings and classrooms. Although there is consensus that the evidence base linking school infrastructure improvements to learning is weak, a number of recent studies have shown impacts on learning. There is also evidence that school infrastructure improvements contribute to increased access and attendance, a stepping-stone toward learning.

Evidence suggests that improvements in school and classroom infrastructure can positively impact student enrollment and achievement in developing countries. Bagby et al. (2016), Cuesta et al. (2016), and Levy et al. (2009) have suggested that school infrastructure expansions and improvements can have a positive impact on student enrollment and achievement. However, these results may take some time to manifest and are specifically linked to new school (or classroom) construction and adding libraries to schools. These infrastructure improvements induce students to enroll, potentially to stay longer in school and experience a high quality learning environment. Over time, students tend to learn more because they remain in school longer, which manifests in higher test scores in the long term. Thus, for this component to be successful, it is important to view infrastructure improvements and the associated O&M plans as being a long-term reform and capacity-development activity that engages the local community, not simply a short-term fix (Land 2000).

C. Pedagogical improvement and investments in pedagogical innovations

A growing body of evidence indicates that teacher professional development interventions can be effective at improving student learning outcomes. Effectiveness, however, depends on the specific characteristics of the intervention, the context in which they are implemented, and the quality and duration of the intervention. Evans and Popova (2015) examined six reviews of educational interventions in LMICs at both the primary and secondary levels (Conn 2014; Glewwe et al. 2014; Kremer et al. 2013; Krishnaratne et al. 2013; McEwan 2015; Murnane and Ganimian 2014). They found broad support for the effectiveness of pedagogical interventions, long-term teacher training (particularly in-service teacher training programs), and accountability-boosting interventions, all of which are components included in the Secondary Education Activity.

However, Murnane and Ganimian (2014) highlighted substantial impact heterogeneity within these intervention categories. For example, within the category of in-service teacher training, interventions that provide general guidance tend to be ineffective, whereas interventions that tailor support to teachers' skill levels and focus on a specific subject area or skill set are more effective (Popova et al. 2018). The World Bank (2018) also notes that in-service teacher professional development needs to be practical and ensure that pedagogical support personnel give teachers concrete instructions, examples, and even manuals to use for lesson planning. The

World Bank (2018) also noted that in-service training needs to be continuous, occur over a longer period, and build on previous support. For example, holding a one-day workshop for teachers usually does not provide the necessary depth of understanding and is insufficient to affect teacher learning or to change and improve classroom practices in the long term (Boller et al. 2004; Raikes et al. 2006; Winton and McCollum 2008). Pedagogical interventions involving long-term teacher mentoring or in-school teacher coaching tend to produce more sizeable effects on student learning (Conn 2014). Unfortunately, many large-scale government-funded programs do not share many of the characteristics and investments needed to reach success (Popova et al. 2018).

D. School life improvements

Of the main outputs of the Attahadi subactivity, the evidence base is least developed for the impact of school-sponsored extracurricular activities (ECA) on secondary school students.[5] On one hand, a developmental framework suggests that ECA participation confers life skills in the form of a strong work ethic, self-esteem, resilience, and social intelligence, which in turn translate into improvements in academic performance. On the other hand, a zero-sum framework argues that ECA participation diminishes school performance as students spend more time on these activities at the expense of their academic studies (Seow and Pan 2014). The correlational and qualitative evidence from developing countries largely support the positive contribution of ECA participation on soft-skills; studies have associated organized sports in particular with better social-emotional skills such as self-respect, confidence, self-esteem, self-control, goal setting, communication and expression skills, and an ability to work in teams among participants (Muller Mariano and da Silva Filho 2015; Burnett 2015; Gaible 2015; Khan and Jamil 2017; Maebuta 2011). Further evidence from evaluations of after-school orchestra programs suggest that participants improve on measures of self-control, self-perceptions, and anger (Cid 2017). Many types of ECAs however exist and studies currently offer little guidance on the programmatic or contextual factors that might drive impacts.

E. Student skills, employment outcomes, and earnings

Educational programs can increase individual wages when they collaborate with the private sector and expose youth to the types of skills employers seek as well as to opportunities to gain experience through internships and job shadowing (Ibarraran et al. 2014). In addition to literacy, numeracy, and other technical skills, soft skills have been posited to play a vital role in a range of life outcomes, including productivity and employment. Soft skills refer to a broad set of skills, personality traits, and personal qualities that enable people to effectively navigate their environment (Gates et al. 2016). Kautz et al. (2014) suggest that soft skills could rival IQ in predicting educational attainment, success in the labor market, health, and even criminality. However, there is currently a lack of rigorous evidence on the effectiveness of soft skills training to improve economic outcomes, especially in LMICs (Blattman and Ralston 2015; Rankin et al. 2015). Although existing research shows that educational interventions can improve soft skills

[5] "School life" in Morocco is understood in general to refer to extracurricular activities in school.

(Durlak et al. 2011), there is little research on the association between soft skills training and earnings and employability. The literature also shows mixed results on improved economic outcomes for youth employability programs that are focused on soft skills (Groh et al. 2016).

F. Country-specific and international relevance of the evaluation

The evaluation of the Secondary Education Activity in Morocco has the potential to contribute to gaps in the literature on how and why specific education interventions that include school-based management, school infrastructure, pedagogical training, and the use of data for decision making can result in students' improved educational attainment, learning of technical and soft skills, and employability. A rigorous evaluation of the Attahadi model package of activities, which will be complemented by an implementation study of the Attahadi model, MASSAR, and O&M subactivities, also has the potential to make important contributions to policymaking in Morocco in four ways. First, the studies will allow MCC and the GoM to rigorously attribute effects to the package of interventions in the Attahadi model on key education outcomes and to gain a deeper understanding of how institutional changes at the school, regional, and national levels affect the key outcomes of interest. Second, this evaluation can provide evidence on whether and how improving secondary school infrastructure and school management leads to students staying in school and learning more. Third, these studies have the potential to make substantive contributions to policy and knowledge in soft skills acquisition. Finally, our studies will help policymakers and educators understand and effectively use data for decision making at all levels to improve education in Morocco. The studies will help GoM and MCC understand both the facilitators to use of data and the barriers that prevent its use, and provide insights into how trained staff have used information effectively to help improve the education system at all levels.

IV. Evaluation Design

The goals of the evaluation of the Secondary Education Activity are to estimate the impacts of the Attahadi model on student, teacher, and school outcomes and to describe and understand the implementation process for all three subactivities. We will use a mixed-methods approach with two components: (1) an RCT of the school-based subactivity, including a cost analysis, and (2) an implementation study of all three subactivities. This report focuses on the RCT evaluation of the Attahadi model.

Table IV.1 summarizes the evaluation questions and the method we will use to answer each question. The questions are subdivided by subactivity and by the unit of measurement (student, teacher, school, or system). This evaluation does not address questions about the impact of the interventions on long-term outcomes like employment, as measuring change in such outcomes would require a longer time horizon. Instead, we focus on measuring change in medium-term outcomes (student completion and learning), which can be reasonably expected during the time frame for this evaluation. These intermediate outcomes are critical foundational indicators to better future employability so we believe the results will provide insights into potential changes in individual and firm productivity in the future (Abarcar et al. 2018).

Table IV.1. Evaluation questions and design

		RCT	Implementation study
Attahadi subactivity			
Student	**1.** What are the impacts on learning (numeracy, literacy, and soft skills)?	X	-
	2. What are the impacts on key educational outcomes, including enrollment, completion, and attendance?	X	-
	3. Are there differential impacts by gender across educational outcomes?	X	-
Teacher	**4.** What are the impacts on teaching and how were the impacts obtained?	X	X
	5. What are the impacts on teachers' attendance?	X	-
	6. Did the Attahadi interventions improve school management and lead to improved accountability among teachers? If so, how?	-	X
School	**7.** Have institutional autonomy and accountability manifested in the participating schools? If yes, how have these things manifested themselves?	-	X
	8. What are the impacts on the quality of infrastructure and physical environment of the school?	X	-
	9. How did the size of the budget managed by schools and the common uses of this budget change?	-	X
	10. How is the decentralization process being incorporated in schools?	-	X

		RCT	Implementation study
Assessment and EMIS subactivity			
System	**11.** How did the interventions contribute to improved student assessment, data, and policy feedback in the EMIS system (MASSAR), leading to a more performance-driven education system?	-	X
O&M subactivity			
System	**12.** How do the infrastructure improvements and new O&M plan lead to an improved and sustainable learning environment?	-	X
Overall sustainability			
System	**13.** To what extent can the Moroccan Ministry of Education sustain the interventions under the Secondary Education Activity?	-	X
	14. To what extent are the interventions under the Secondary Education Activity cost-effective? (i.e., can the Ministry of Education financially sustain the interventions? What is the economic rate of return to the beneficiaries?)	X	-

A. Methodology

The impact evaluation uses an RCT to estimate the causal impacts of the Attahadi subactivity. RCTs consist of randomly assigning units (in this case, schools) to either a group that receives the intervention (the treatment group) or one that does not (the control group). Random assignment ensures that school, teacher, and student characteristics do not determine treatment status and as a result, characteristics should be the same on average in treatment and control groups prior to the intervention.[6] Thus, the control group represents what would have happened to the treatment group in the absence of the intervention. Comparing the outcomes between the treatment and control groups after exposure to the intervention will provide the causal impact of the program.

MCC, MCA-M, and the Ministry of National Education, Vocational Training, Higher Education and Scientific Research (MENFPESRS) selected 3 out of 12 regions that were seen as nationally representative (MCC 2015a).[7] Four provinces were selected within each of the three regions (Table IV.2). In each province, MCA-Morocco and the MENFPESRS screened all lower secondary and upper secondary schools for eligibility to participate in the activity. The selection criteria for eligible schools included schools that (1) were not condemnable or slated for demolition, (2) did not have asbestos, (3) were not undergoing rehabilitation (or had not yet reopened after a rehabilitation), (4) did not have major structural problems, (5) had no ongoing legal case against the school, such as a property claim or dispute, and (6) had at least 288 students enrolled (50 percent of the built capacity of the smallest model school) (MCC 2015a).

[6] Any differences between treatment and control groups at baseline would have arisen due to chance.

[7] In order to achieve national representation to maximize the relevance of learning for post-compact scale-up, MCC and the GoM chose the three regions to balance the following four criteria: (1) representation of the northern, central, and southern regions of the country; (2) strength of economic and job growth potential; (3) poverty rates; and (4) representation of high, medium, and low educational outcomes.

In each region, the MENFPESRS and MCA-M determined the eligibility of schools to participate in random assignment through site inspections and verification of school records. Through this process, 306 lower secondary (LS) and 171 upper secondary (US) schools were determined to be eligible for random assignment across all three regions. These numbers exclude 6 schools (3 LS and 3 US) that were preselected to receive the pilot program in TTH and therefore did not participate in random assignment and are included in the RCT. Mathematica conducted random assignment of schools in the TTH region in December 2016 and in FM/MS in March 2018. We stratified schools by province, urban or rural status, and school type (whether LS or US) to ensure balance on those key characteristics and to improve the precision of our estimates of impacts on student learning.

Once the selection of eligible schools was completed, Mathematica, in collaboration with MCC, MCA-M, and the MENFPESRS, proceeded with random assignment. The process involved three steps:

1. **Identifying strata**. The first step was to group the schools by strata**.** We divided schools according to their province, urban or rural status, and school type (whether LS or US).
2. **Determining the number of schools to select by strata**. The second step was to calculate the number of schools to select for the treatment group in each stratum. Of the 28 schools selected to receive the Attahadi model through random assignment, MCC, MCA-M, and the MENFPESRS determined that a total of 19 LS schools and 9 US schools would be selected each in Tanger-Tétouan-Al Hoceima and Marrakech-Safi and 18 LS schools and 10 US schools would be selected in Fès-Meknès. The number of treatment schools selected for each stratum was determined in proportion to the number of eligible schools in each stratum. This selection process gave schools in each stratum an approximately equal chance of selection for the Attahadi model.
3. **Conducting random assignment through public lottery**. The final step was to hold a public lottery for random assignment for each of the three regions. The lottery occurred for schools in the TTH region in December 2016 and in FM/MS in March 2018. MCA-M and Mathematica organized a public ceremony with central, regional, and provincial level representatives of the MENFPESRS to ensure that the random assignment process was transparent in all regions. We conducted random assignment by drawing wooden blocks from bags. School authorities, government officials and students took turns selecting the requisite number of wooden blocks from each bag. The schools chosen in the ceremonies (the treatment group) received the Attahadi model interventions while the rest (control group) continued with business as usual. We note that the ceremonies were well-attended by stakeholders and no complications or procedural issues during the events that could have compromised random assignment.

Table IV.2. Number of eligible and selected schools by region, province, and stratum

Region	Province	Eligible schools LS schools / US schools			Selected schools LS schools / US schools		
		Urban	Rural	Urban/ rural	Urban	Rural	Urban/ rural
Tanger-Tétouan-Al Hoceima	Chefchaouen	3	9	10	1	2	2
Tanger-Tétouan-Al Hoceima	Larache	13	7	8	3	1	2

Region	Province	Eligible schools LS schools / US schools			Selected schools LS schools / US schools		
		Urban	Rural	Urban/ rural	Urban	Rural	Urban/ rural
Tanger-Tétouan-Al Hoceima	Tanger	28	0	17	6	0	3
Tanger-Tétouan-Al Hoceima	Tétouan	17	8	12	4	2	2
Fès-Meknès	Fès	46	1	28	7	0	4
Fès-Meknès	Ifrane	4	7	5	1	1	1
Fès-Meknès	Meknès	32	7	21	5	1	3
Fès-Meknès	Taounate	7	14	13	1	2	2
Marrakech-Safi	Chichaoua	3	10	9	1	2	1
Marrakech-Safi	Essaouira	5	13	8	1	3	1
Marrakech-Safi	Marrakech	34	13	30	6	2	5
Marrakech-Safi	Safi	12	13	10	2	2	2
Total		**204**	**102**	**171**	**38**	**18**	**28**

Note: LS = lower secondary; US = upper secondary. Urban and rural upper secondary schools within each province were combined into the same stratum. Schools that had both lower secondary and upper secondary levels on the same premises (31 out of 477 schools) were considered upper secondary schools for random assignment and for the evaluation.

B. Impact evaluation study sample

The evaluation study population consists of all schools that participated in random assignment and were eligible to receive the Attahadi model. This includes all students and teachers in these schools. We obtained administrative data from the MASSAR system to examine the characteristics and outcomes of this population. We also gathered more detailed survey data on a representative subsample of schools, teachers, and students to be able to answer all our research questions.

The following describes our sampling procedures for baseline survey data collection on schools, teachers, and students. MCA-M engaged a local service provider to collect reliable, high quality data.

1. School sample

We collected survey data from all 84 treatment schools and from a random subsample of 84 schools out of all control schools not selected during random assignment. We chose an equal number of control schools for the survey sample to balance statistical power for the evaluation and the cost of data collection. This process resulted in a total of 168 schools in the sample.

2. Teacher sample

In each of the 168 sampled schools, we randomly selected a sample of six teachers who were the target of the planned pedagogical training to participate in the survey. To develop the sampling frame, we obtained administrative data on all teachers and the subjects they taught and worked with the MENFPESRS to identify the teachers who were most likely to benefit from the training. In the control schools, the sampling frame consisted of teachers who would have received

training had their school been selected to receive the Attahadi model.[8] We also randomly sampled four additional teachers to serve as replacements if teachers were unable or unwilling to respond to the survey at the time of our school visits.

3. Student sample

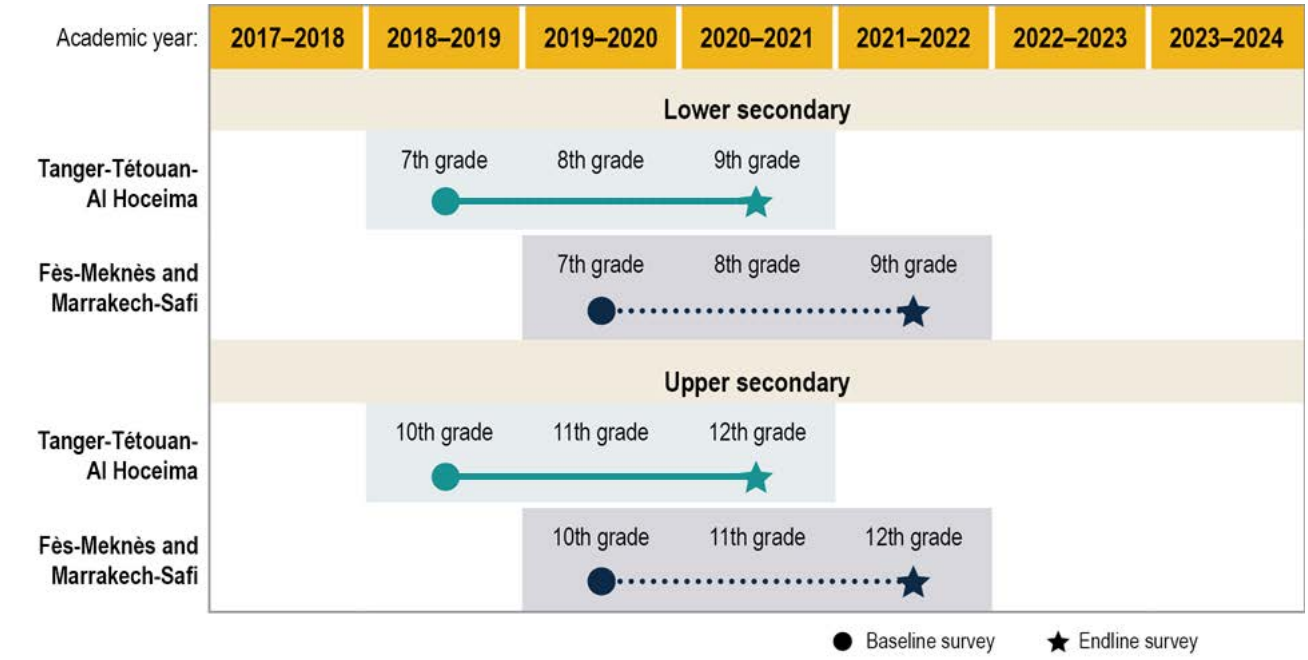
To create the student survey sampling frame, we obtained administrative data on all students in the 168 schools. We randomly sampled 15 7th grade students in lower secondary schools and 15 10th grade students in upper secondary schools. Similar to the teacher sample, we also randomly sampled 10 additional students per grade to serve as replacements.

Based on the original Attahadi model implementation plan, we initially collected baseline data from TTH students entering the target grades in the 2017–2018 school year. However, owing to implementation delays, those students would have been exposed to the activities for less time than expected. To maximize the time of exposure to program activities, we decided to focus on the cohorts entering school one year later in each region and therefore collected another baseline from the TTH cohort of students in the target grades during the 2018–2019 school year. We used the same student survey instrument in both baselines.

This report focuses on the baseline data collected in TTH in May 2019, from students who were in 7th and 10th grades in the 2018–2019 school year; and in FM/MS in November 2019, from students who were in the target grades in the 2019–2020 school year. The original plan was to collect endline data from the same students in April/May 2021 in TTH and April/May 2022 in FM/MS after two years of exposure (Figure IV.1), but given further implementation delays exacerbated by COVID-19 pandemic restrictions in Morocco, these cohorts may no longer be exposed to the full package of interventions for two years. As a result, we have decided to delay endline data collection for these students and are actively considering alternative plans, including changing the student cohorts that we will follow for the evaluation. We are consulting with MCC about the options and will update our Evaluation Design Report, as appropriate.

[8] These include teachers who taught the following subjects: history and geography, philosophy, physics and chemistry, natural science, mathematics, English, Arabic, and French.

Figure IV.1. Student cohorts followed across time



Note: LS= lower secondary; US = upper secondary. The lines represent the longitudinal sample of students whom we intend to follow for the evaluation. These are the 7th and 10th grade students in Tanger-Tétouan-Al Hoceima in the 2018–2019 school year (represented by the solid green lines) and the 7th and 10th grade students in Fès-Meknès and Marrakech-Safi in 2019–2020 (respresented by the dotted blue lines). The circles show when we will conduct the baseline survey and the stars indicate when we will conduct the endline survey of these students.

V. Data Sources, Outcome Definitions, and Analytic Approach

A. Overview of baseline data sources and outcomes

This section provides an overview of the primary and secondary data sources used in this baseline report. Primary data collection included surveys with students, teachers, and school directors, and a school infrastructure observation checklist for each school. Secondary data consisted of administrative data from the MASSAR system and school directors. We use these data to describe the evaluation sample and assess baseline equivalence on the outcome variables.

Survey (n = 2,520 students, 1,008 teachers, 168 directors). We conducted in-person surveys with students, teachers, and school directors. We surveyed six teachers, 15 students (7th graders for lower secondary schools and 10th graders for upper secondary schools), and the school director at each school in the sample.

School infrastructure checklist (n = 168 schools). Enumerators visually assessed the quality of the school's infrastructure by completing a school infrastructure checklist during the visit to the school. This checklist gathers information on the existence and quality of the school infrastructure.

School administrative data (n = 168 schools). Enumerators collected school administrative data from school records in all schools. The information was collected in paper form as part of the survey that the enumerators applied to school directors. This data source provides information on the school budget, student enrollment, and teacher and student absenteeism.

Administrative MASSAR data (n = over 400,000 students, 17,000 teachers, and 400 schools in each school year). We obtained end-of-school-year data (June) for all 7th and 12th grade students, teachers, and schools in the three study regions for the 2017–2018 and 2018–2019 school years. The data capture end-of-school-year outcomes, such as individual test scores and progression and dropout for students, and absenteeism for students and teachers. These data include school and student codes that allow us to link back the administrative data to treatment and control schools and the survey sample.

Table V.1 presents a description of the primary outcomes in this report from each of these data sources.

Table V.1. Baseline outcomes and data sources

Domain	Outcome description	Data source
Student outcomes		
Academic performance	Student scores on the **local, regional, and national exams** and **students' passing grade (moyenne passage)**. Lower secondary students take the local and regional exams at the end of 9th grade. A student's overall passing grade (moyenne passage), determines their completion of lower secondary school. The moyenne passage is computed as a weighted average of the local exam score (30 percent), regional exam score (40 percent), and continuous assessments from both semesters of 9th grade (30 percent). Similarly, upper secondary students take the regional exam in 11th grade and the national exam in 12th grade. A students' overall passing grade (moyenne passage), determines their completion of upper secondary school. The moyenne passage is computed as a weighted average of the regional exam score (25 percent), national exam score (50 percent), and continuous assessments from both semesters of 12th grade (25 percent).	MASSAR administrative data (each exam is scored out of 20 points)
Soft skills[a]	**Openness.** Consists of traits such as having wide interests, being open to new things, coming up with new ideas, valuing aesthetic and artistic experiences, and having an active imagination.	Self-report Adapted version of the Big Five Inventory scale (Plaisant et al. 2010) (45 items)
	Conscientiousness. Encompasses traits such as reliability, efficiency, organization and thoroughness, and making plans and following through with them.	
	Extroversion. Includes traits such as being talkative, energetic, enthusiastic, outgoing, and assertive.	
	Agreeableness. Refers to how much students value getting along with others. It includes traits like unselfishness, trust, and being kind, forgiving, and affectionate.	
	Neuroticism. Tendency to experience negative emotions. It includes traits such as being tense, nervous, moody, and worrying excessively.	
	Grit. Also referred to as goal orientation, it refers to perseverance and resilience and manifests in traits such as working hard, working on tasks that take a long time, completing tasks, and willingness to fail and try again at a task.	Self-report on an adapted version of the GRIT scale (Duckworth et al. 2007) (5 items)
Attachment to school	**Absenteeism.** Number of days student was absent from school in the past week.	Self-report
	Absenteeism. Number of days student was absent from school in the past month, with and without justification.	School administrative records
	Dropout. Percentage of students who dropped out during or between school years.	MASSAR administrative data
Academic attainment	**Progression.** Percentage of students in a given grade during the 2017–2018 school year who enroll in the next grade level for the 2018–2019 school year (TTH, FM, and MS).	MASSAR administrative data
	Repetition. Percentage of students repeating the grade level for the 2018–2019 school year (TTH, FM, and MS).	MASSAR administrative data
	Graduation. Percentage of 9th and 12th grade students graduating at the end of the 2017–18 school year (TTH) or the end of the 2018–19 school year (FM and MS).	School administrative records

Domain	Outcome description	Data source
Teacher outcomes		
Teaching methods	**Pedagogical knowledge and attitudes**. Techniques and strategies that teachers use to help students learn in the classroom, such as organizing group topic discussions, soliciting student input, using interactive lecturing, etc.	Self-report 18 items
Learning environment	**Time spent in school**. Number of hours teachers spend in school on a typical day.	Self-report
	Time use for school activities. Number of hours, at home or in school, teachers spend per week on school activities, such as planning and preparing lessons, grading tests, meeting with parents, providing academic support to students, etc.	Self-report (11 items)
	Absenteeism. Number of days teacher was absent from school in the past week.	Self-report
	Absenteeism. Number of days teacher was absent from school in the past month, with and without justification.	School administrative records
School outcomes		
School outcomes	Condition of the main school building. Whether the building or item is sound, requires some preventive or curative repairs, or requires significant repairs to function. This also includes the presence of an enclosure, wheelchair access ramps, and a separate space for faculty/administrative personnel.	School infrastructure observation checklist (8 items)
	Condition of the school classrooms. Students have individual desks, classroom temperature, and a 16-item classroom checklist (entrance door, entrance door is at least one meter wide, entrance door opens and closes, entrance door locks, functioning electric lights, at least one window, black or white board, board can been seen from the back of the classroom, student tables, student tables are not bolted down, student desks, student desks are not bolted down, student chairs, student chairs are mobile, teacher desk, and platform).	School infrastructure observation checklist (18 items)
	Condition of water, sanitation, and hygiene (WASH) facilities. Separate toilets for boys and girls, flush toilet sewer connection, Seven-item checklist for boys and girls toilet facility conditions (handicapped-accessible, interior doors for toilet stalls, functioning door, running water, soap, electric lighting, at least one window), toilet facilities inside the school, and number of toilet stalls available.	School infrastructure observation checklist (20 items)

[a] We will work with local stakeholders to identify appropriate measures of soft skills for endline data collection.

B. Survey baseline data collection response rates

We collected survey data from all 168 schools in the evaluation sample. We surveyed both treatment and control schools at the same time in each region to prevent any issues related to differential timing of data collection between treatment and control schools. In each school, we randomly chose the teachers and students to include in the sample using administrative data. We replaced individuals in the surveys if they were chosen but refused to participate in the study, were absent on the day of data collection, or were no longer employed or enrolled at school. Enumerators were given a replacement list of randomly chosen students and teachers that Mathematica prepared before data collection began. By sampling with replacement, we obtained our target sample sizes for the full sample.

All school directors in the 168 schools agreed to participate in the survey, which resulted in a 100 percent response rate. Over 80 percent of teachers originally chosen to be included in the study sample responded to the surveys. The corresponding response rate for students in the initial

sample was over 60 percent. The low response rates (high replacement rates) associated with the sample of students and teachers is primarily related to absenteeism or the students' or teachers' schedules, which accounted for over 70 percent of nonresponses in both teacher and student samples. Other reasons for nonresponse include refusal to participate in the survey or because individuals were no longer enrolled or working at the school. Response rates of the initially sampled teachers and students do not differ significantly between treatment and control schools, which suggests that selection into the survey sample should not affect the results of our comparisons between the treatment and control groups. Table V.2 shows these response rates for lower and upper secondary teachers and students in treatment and control schools.

Table V.2. Baseline data collection response rates for teachers and students originally chosen to be included in the study sample

	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (c)
Teachers				
Lower secondary teachers	85.4	86.9	-1.5	0.563
Upper secondary teachers	86.9	82.7	4.2	0.238
Students				
Lower secondary students	60.1	59.9	0.2	0.881
Upper secondary students	63.8	63.1	0.7	0.534

Note: All values shown are percentages. Response rates refer to the percentage of surveys that were completed by individuals who were originally chosen for the study sample (i.e. individuals who did not need to be replaced).

C. Analytical approach

To assess differences between intervention and control groups of schools, teachers, and students, we conducted regression analyses that adjust for strata that were used during random assignment. We estimated school- and individual-level (student and teacher) regressions.

1. Estimation approach

For each characteristic or baseline outcome examined, the general regression model we use to test for differences between study groups is as follows:

$$(1)\quad Y_{i,j} = \alpha + \beta_1 T_j + \delta_k + e_{i,j}$$

where $Y_{i,j}$ is the characteristic or outcome of interest for student or teacher *i* in school *j* at baseline and T_j is an indicator for treatment, equal to 0 for students or teachers in schools assigned to the control group and 1 for students or teachers in schools assigned to the intervention. The coefficient of interest is the parameter β_1, which captures the difference between the intervention and control group in each characteristic or outcome and its statistical significance. δ_k represents a vector of indicator variables for the strata within which random

assignment of schools was conducted. Finally, e_{ij} is a teacher- or student-level error term.

Because random assignment was conducted at the school level, the teacher and student regressions use clustered standard errors to account for the nesting of students and teachers in schools. For school-level estimates, no clustering adjustment was necessary because all terms in Equation (1) are at the school level.

Equation (1) was estimated separately for the sample in lower secondary and upper secondary schools. Schools that have both lower and upper secondary levels were counted as upper secondary schools for the purpose of random assignment, student sampling, and analysis.[9]

2. Strata

All our regression models include stratum dummies (δ_k) to account for our random assignment approach. Stratification (conducting random assignment within groups of similar schools) improves efficiency by reducing the amount of variation between schools of different treatment status.

During random assignment, we created 36 strata based on schools' province, urban or rural status, and school type (LS or US). It was necessary to combine urban and rural US schools into the same stratum within each province because there were only a few rural schools in these strata with which to conduct random assignment. In addition, for those schools that had both lower and upper secondary levels on the same premises—which, therefore, would both be affected by the intervention activities—we consulted with stakeholders and agreed to consider these schools as US schools for the purposes of random assignment, sample selection, and analysis.

[9] 31 out of 477 schools at random assignment had both lower and upper secondary levels in the same premises. 10 out 168 schools in the survey sample had both levels.

This page has been left blank for double-sided copying.

VI. Findings

In this chapter, we present baseline findings for the Morocco Secondary Education evaluation. First, we discuss the external validity of the evaluation and present findings of descriptive statistics of key variables, comparing the data collected through our survey to data on the population to assess the data's external validity. Second, we conduct balance tests between the treatment and control groups to assess their equivalency. Third, we assess the equivalency of the two groups by student gender as a goal of the evaluation is to assess the impacts of the intervention by gender. The results of our analysis show that, overall, our study sample represents the broader population of students, teachers, and schools that are the target of the Attahadi model and that the treatment and control group schools are equivalent, with a few minor differences that we will control for statistically using covariate adjustment in our follow up analyses at endline.

A. External validity

We designed the impact evaluation so that our survey findings would apply to the broad group of students, teachers, and schools eligible to receive the Attahadi model (i.e. all participants in the lotteries) in the three study regions. We randomly sampled students, teachers, and schools for the surveys to make the survey data representative of this population. Random sampling ensures that findings from the survey data would apply to Attahadi model beneficiaries and other similar beneficiaries should the Moroccan government decide to expand the interventions to more schools in the three regions. In addition, because the three regions were selected to be nationally representative, the findings should be relevant should the government decide to expand to schools outside of the three regions.

In this section, we assess external validity by comparing the characteristics of our TTH sample to the Attahadi model eligible population in the same region using results from our surveys and administrative data from the MENFPESRS. If the survey sample possesses similar characteristics to the broader population, then this increases our confidence that evaluation results based on the survey data—where we collect more detailed information on outcomes than in the administrative data—would generalize to the wider group. For now, we focus on the TTH region because we do not yet have the population data for FM and MS.[10] At endline, we will perform the same analysis for the FM and MS to further assess generalizability in these two regions.

In TTH, the sample of 7th and 10th graders resembles the broader student population in the same grade levels (Table VI.1). The student sample consists of a similar percentage of females (45 percent for lower secondary and 59 percent for upper secondary students) as the population in the region. Lower secondary students in the sample and population are also the same age (mean age 13.7 years). Although upper secondary students in the sample appear

[10]We obtained MASSAR data for all regions for the 2017-2018 and 2018-2019 cohorts but have yet to obtain the 2019-2020 data. The survey data for the TTH cohorts are for 2018-2019 while the survey data for the FM/MS cohorts are for 2019-2020. Thus, we could only make comparisons for the TTH cohorts at this point.

slightly younger than students in general (mean age 16.5 versus 16.7 years), the estimated difference is small, though statistically significant.

Table VI.1. Comparison of administrative population and survey sample—students

	Population (A)	Sample (B)	Difference (A-B)	*p*-value (C)
Lower secondary students				
Age	13.8	13.7	0.1	0.645
Female (percentage)	46.0	45.2	0.8	0.722
Sample size lower secondary students	**27,756**	**570**	**-**	**-**
Sample size lower secondary schools	**85**	**38**	**-**	**-**
Upper secondary students				
Age	16.7	16.5	0.2**	0.027
Female (percentage)	56.5	58.7	-2.2	0.463
Sample size upper secondary students	**14,253**	**254**	**-**	**-**
Sample size upper secondary schools	**45**	**17**	**-**	**-**

Source: MASSAR data for the 2018–19 school year (TTH).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

The sample of teachers in the study generally resembles the teacher population in TTH but is slightly different on several dimensions, including age, experience level, and the percentage of female teachers in upper secondary schools (Table VI.2). Our comparisons focus on the broader population of teachers who teach history and geography, philosophy, physics and chemistry, natural science, mathematics, English, Arabic, or French—the same as our evaluation sample. On average, lower secondary teachers in the evaluation sample are two years older and have two additional years of teaching experience compared to the broader teacher population. These differences are statistically significant at the 5 percent level. Teachers' level of education is largely similar between the sample and population, except for two categories: lower secondary teachers in the sample were more likely to attain a master's degree (by four percentage points) signifying that the sample is slightly more educated than the population, whereas upper secondary teachers were less likely to report educational attainment in a non-specified category (by one percentage point). There is also a much lower percentage of female teachers in upper secondary schools for the sample than in the population (33.6 versus 41.2 percent).

We are unclear why such differences exist given that we employed random sampling and have no reason to believe that the selection process was compromised when conducting the surveys. However, we see several possible explanations. First, some differences could simply be due to chance from random sampling. We conducted 26 hypothesis tests and expect to find about 3 statistically significant differences by chance (at the 10 percent significance level) from random variation. Second, the differences may reflect differences in measurement between the survey

and administrative datasets. For example, years of employment in the administrative data are documented based on actual start dates, while our survey data used teacher recall to collect this information. It is possible that teachers in our survey may have overestimated the number of years they had been teaching when asked in the survey. Third, differential response rates among respondents with particular characteristics could have played a role in creating the differences that we see in the results. For example, female teachers from upper secondary schools in TTH were 15 percentage points (p-value = 0.090) less likely to complete the survey compared to male teachers.[11] This result, in addition to our relatively small sample size for teachers, could have translated to the difference detected between the sample and population in gender for upper secondary teachers. We have no way to test these explanations.

Nevertheless, while the result of this analysis leaves us less confident about the external validity of results we will obtain from the teacher survey sample (relative to the student and school samples) the differences between the sample and population are not large in magnitude, except perhaps for the gender of upper secondary school teachers. At endline, we plan to exert more effort in achieving high follow-up response rates for the teacher survey to avoid reducing the representativeness of the teacher survey sample further.

[11] As discussed, we sampled with replacement, but replacements were randomly selected and did not necessarily have the same characteristics as the respondent who could not complete the survey (i.e. we did not necessarily replace female teachers who could not complete the survey with female teachers in the sample).

Table VI.2. Comparison of administrative population and survey sample—teachers

	Population (A)	Sample (B)	Difference (A-B)	*p*-value (C)
Lower secondary teachers				
Teacher age	40.5	42.6	-2.2***	0.005
Female (percentage)	53.0	52.8	0.2	0.956
Years of teaching in the school	8.0	9.2	-1.2	0.102
Years of teaching in total	14.1	16.2	-2.2**	0.011
Highest level of education (percentage)				
Upper secondary	14.2	16.4	-2.2	0.346
First university cycle (DEUG)[a]	9.5	11.6	-2.1	0.276
Second university cycle (*licence*)	65.5	65.1	0.4	0.900
Third university cycle (master's)	9.8	6.0	3.9***	0.008
Doctorate	0.4	0.5	-0.1	0.915
Other	0.5	0.4	0.1	0.888
Teacher position (percentage)				
Permanent	84.2	83.7	0.5	0.838
Contract	15.8	16.3	-0.5	0.838
Other	0.0	0.0	0.0	1.000
Sample size lower secondary teachers	**1,885**	**211**	**-**	**-**
Sample size lower secondary schools	**85**	**45**	**-**	**-**
Upper secondary teachers				
Teacher age	41.6	43.2	-1.6	0.232
Female (percentage)	41.2	33.6	7.6*	0.073
Years of teaching in the school	8.0	8.3	-0.3	0.759
Years of teaching in total	15.0	16.3	-1.3	0.366
Highest level of education (percentage)				
Upper secondary	6.0	4.5	1.5	0.379
First university cycle (DEUG)[a]	2.0	3.1	-1.2	0.448
Second university cycle *(licence*)	57.3	60.3	-3.0	0.522
Third university cycle (master's)	29.5	26.9	2.6	0.616
Doctorate	4.0	5.1	-1.1	0.621
Other	1.3	0.0	1.3***	0.000
Teacher position (percentage)				
Permanent	93.4	96.2	2.1	0.188
Contract	6.6	3.8	2.1	0.188
Other	0.0	0.0	0.0	1.000
Sample size upper secondary teachers	**2,071**	**103**	**-**	**-**
Sample size upper secondary schools	**47**	**22**	**-**	**-**

Source: MASSAR data for the 2017–18 school year (TTH).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.
[a] DEUG = Diplôme d'études universitaires générales.

The school and school director characteristics from both the sample and the broader population in TTH are similar. We found only two significant differences (out of 26 comparisons) when assessing results for both groups (Table VI.3). The differences pertain to the highest level of education of school directors. Lower secondary school directors are more likely to have obtained a DEUG than a *licence*. Overall, the results give us confidence that the impact estimates for the sample are likely representative and could apply to the broader group of potential beneficiaries at the school level.

Table VI.3. Comparison of administrative population and survey sample—schools and school directors

	Population (A)	Sample (B)	Difference (A-B)	*p*-value (C)
Lower secondary schools				
Number of students enrolled	871.3	885.9	-14.6	0.747
Percentage of enrolled students who are female	48.5	48.3	0.1	0.735
Number of teachers employed	31.0	32.0	-1.0	0.410
Percentage of employed teachers who are female	47.1	47.6	-0.5	0.732
School located in rural area (percentage)	27.6	27.6	-0.0	0.426
School director				
Female (percentage)	9.5	5.6	3.9	0.288
Length of tenure as director (years)	7.8	8.0	-0.2	0.838
Highest level of education (percentage)				
Upper secondary	13.4	13.2	0.1	0.976
First university cycle (DEUG)[a]	4.9	-0.2	5.0**	0.044
Second university cycle (*licence*)	73.1	82.1	-9.0*	0.085
Third university cycle (master's)	8.7	4.9	3.8	0.224
Doctorate	0.0	0.0	0.0	1.000
Other	0.0	0.0	0.0	1.000
Sample size lower secondary schools	**85**	**38**	**-**	**-**
Upper secondary schools				
Number of students enrolled	1186.8	1160.7	26.1	0.781
Percentage of enrolled students who are female	53.7	52.9	0.8	0.592
Number of teachers employed	58.4	60.1	-1.6	0.696
Percentage of employed teachers who are female	38.7	38.0	0.7	0.643
School located in a rural area (percentage)	21.7	15.7	6.0	0.253
School director				
Female (percentage)	4.3	6.0	-1.7	0.682
Length of tenure as director (years)	5.9	6.2	-0.3	0.745
Highest level of education (percentage)				
Upper secondary	13.6	15.9	-2.3	0.700
First university cycle (DEUG)	4.2	6.1	-1.9	0.666
Second university cycle (*licence*)	66.4	67.2	-0.8	0.928
Third university cycle (master's)	8.9	5.5	3.4	0.435
Doctorate	6.8	5.3	1.5	0.707
Other	0.0	0.0	0.0	1.000
Sample size upper secondary schools	**47**	**18**	**-**	**-**

Source: MASSAR data for the 2017–18 school year (TTH).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. This may result in small negative percentages. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] DEUG = Diplôme d'études universitaires générales.

Although the overall results show that the survey sample and population have similar characteristics, caution is still necessary when extending the findings to any of the three regions. There are two main caveats to this analysis. First, we were only able to assess the representativeness of the TTH cohorts, due to data availability. We have not obtained the 2019–2020 MASSAR administrative data to use to compare the FM/MS evaluation samples (that were drawn in that school year) with the population. Second, impact estimates for the treatment group may still not be externally valid if the features of training, the broader educational landscape, or the overall economic situation in Morocco changes substantially in the future.

B. Balance tests for treatment and control groups

In this section, we present balance tests to assess the equivalence of the treatment and control groups and determine if the two groups are similar in observable characteristics at baseline. This would provide evidence that randomization succeeded in creating study groups that are similar in all baseline characteristics (both in observed and unobserved characteristics) except for receipt of the Attahadi model We present our findings separately for lower and upper secondary schools. First, we compare baseline demographic and other characteristics between the groups of students, teachers, and schools. Second, we present balance tests for our primary outcomes of interest—students' grit, test scores, and absenteeism; teachers' pedagogical knowledge and absenteeism; and schools' infrastructure characteristics and management. Finally, we provide subgroup analysis of these primary outcomes for students by gender. Overall, our results show that the treatment and control groups are similar in characteristics and outcomes at baseline. We detect some statistically significant differences, but these are likely due to chance, given our randomized design. Nonetheless, the magnitudes of the differences are relatively small and can be statistically adjusted in our endline analysis. Additional findings on the baseline equivalence of the two groups for secondary characteristics and outcomes of interest can be found in Appendix B.

1. Demographic and academic characteristics of students, teachers, and schools

This section describes the evaluation sample and assesses whether the demographic and academic characteristics of students, teachers, and schools are equivalent between treatment and control groups at baseline. We use the baseline survey data collected for the impact evaluation as our primary data source unless otherwise noted in the report. Our analysis finds that upper secondary students, lower secondary schools, and upper secondary schools have equivalent baseline characteristics. There are some statistically significant differences between lower secondary students and lower and upper secondary teachers in both groups; however, these differences are small in magnitude. The full results for these and additional baseline characteristics for students, teachers, and schools can be found in Appendix B, Tables B.1–B.6.

Student demographic and academic characteristics among upper secondary school students are equivalent (Appendix B, Table B.1). These characteristics include age, gender, parents' education, language spoken at home, whether they attended the same school the

previous year, and which academic track[12] they are studying (for upper secondary students). The average student in both groups is 16 years old and the majority speak primarily Arabic at home (96 and 92 percent in treatment and control schools, respectively). Most students are pursuing the general education track (95 percent in treatment and 97 percent in control schools). Fifty-two percent of treatment and 51 percent of control students are female.

We see three statistically significant differences out of 23 comparisons between upper secondary students in the treatment and control groups. These differences occur in students' maternal educational attainment and relate to mothers whose highest level of education was reported as "none," "lower secondary school," or "other."[13] Two of the differences are statistically significant at the 10 percent level and one is significant at the 5 percent level. The differences range from two to eight percentage points and thus are relatively small in magnitude. Overall, treatment students have slightly more-educated mothers than control students.

Lower secondary school students are also similar but have more statistically significant differences than we would expect due to chance (Appendix B, Table B.1). The average lower secondary student in both groups is 13 years old, and 90 percent of students speak primarily Arabic at home. Students' parents have relatively low levels of education. Approximately half of the parents (both mothers and fathers) across the treatment and control groups did not complete education beyond the primary level. In fact, 29 percent of fathers and 50 percent of mothers reportedly never attended school. Nineteen percent of fathers and 14 percent of mothers in the treatment group completed upper secondary or postsecondary school. We have similar findings for control schools (18 percent and 12 percent of fathers and mothers, respectively).

Despite these similarities, lower secondary student demographics differ in three out of 20 characteristics. The differences in student's gender, father's completion of postsecondary school, and whether French is the primary language spoken at home are statistically significant at the 5 and 10 percent levels. However, the differences are relatively small and correspond to less than a five-percentage point difference between the treatment and control groups.

There are several statistically significant differences between teachers in treatment and control schools in terms of their demographic and employment characteristics (Appendix B, Table B.3). The differences occur in characteristics such as years of experience, education, position, track, and whether they teach in multiple schools.[14]

[12] US students have the choice to pursue one of four different courses of study: original, general, technical, and professional. This choice is constrained by the track(s) offered at the student's school. The original education track provides training in Islamic disciplines. The general education track provides scientific, literary, economic, or social training to prepare students for higher education. The technical and professional education tracks provide vocational training.

[13] "Other" was selected when the response did not fall into any of the pre-specified categories, most often because the mother had completed an adult literacy course.

[14] A teacher can be assigned to work in multiple schools simultaneously if there is a need and her weekly hourly workload in a school is below the standard 24 hours for lower secondary schools or 21 hours for upper secondary schools.

Lower secondary teachers in the two groups are similar on most characteristics but differ in terms of their level of education and gender. The teachers have, on average, 15 to 16 years of teaching experience and have been teaching in their current school for an average of 7 years. Over 80 percent of teachers in both groups have a permanent teaching position in the education system and the majority (56 percent of treatment and 63 percent of control) have a bachelor's degree (*license*) as their highest level of education.

There were four statistically significant differences out of 13 comparisons between the groups—which is more than we would expect due to chance. Sixteen percent of teachers in treatment schools have a master's degree, compared to 9 percent of control teachers. This difference of seven percentage points is statistically significant at the 1 percent level. The other differences are statistically significant at the 10 percent level and include gender and having only an upper secondary school diploma or a bachelor's degree.

Upper secondary teachers in the two groups are also similar on most characteristics, but differ across seven characteristics, including years of teaching experience, teaching in multiple schools, employment type, and levels of education. Among upper secondary teachers, about a third are female in both groups. Over 80 percent of teachers have a bachelor's or master's degree and over 85 percent hold a permanent teaching position in the education system. All the teachers in the sample teach the general track and a minority (6–7 percent) also teach the professional or technical tracks. However, we see seven statistically significant differences between the teachers in the treatment and control groups in upper secondary schools, more than would be expected due to chance. Four of these differences are statistically significant at the 5 percent level and the others are significant at the 10 percent level; differences occur for years of teaching experience, whether teachers teach in multiple schools, employment type, and levels of education. These differences are not a product of survey response rates, which are again the same across treatment and control schools (Table V.2). The results suggest that upper secondary teachers in treatment schools are slightly more experienced and educated, on average, than teachers in control schools, but less likely to teach in multiple schools.[15] To adjust for these differences, we will include these variables as covariates in our endline analysis.

Treatment and control schools and school directors in both lower and upper secondary schools are equivalent, with one statistically significant difference between lower secondary treatment and control schools (Appendix B, Table B.5). Schools in both groups have over 800 students enrolled, on average and are in mainly urban areas. Less than 10 percent of school directors are female. School directors have, on average, 14 to 15 years of teaching experience and have been in their position for an average of 8 to 9 years. The only statistically significant difference between the two groups is the percentage of schools that offer boarding to students (significant at the 5 percent level); however, the data also indicate no difference between the

[15] If many teachers teach in both treatment and control schools then it is possible that we do not find any differences in teacher training between treatment and control groups in our final analysis even if teacher training had an effect because many teachers in control schools would have also received training by simply being teachers in treatment schools. However, the threat of this to our evaluation is low given that fewer than 11 percent of teachers teach in multiple schools, and even less teach in both treatment and control schools.

number of students these schools can board, which suggests that the difference in the percentage of schools offering boarding, even if meaningful, are unlikely to affect estimates of student outcomes that are of interest when comparing between study groups. Twenty-five percent of treatment schools offer boarding for students, whereas only 11 percent of control schools offer the same service.

There are no statistically significant differences between upper secondary schools and directors in both groups. The average school in the two groups enrolls just over 1,000 students; most schools are in urban areas and do not offer boarding to students. Very few directors are female, directors have been in their positions for 6 years, on average, and they have an average of 14 to 16 years of teaching experience.

2. Main outcomes of interest

This section presents the results of our analyses on the equivalence of the treatment and control groups as it relates to the primary outcomes of interest, including soft skills, student assessments, absenteeism, and school infrastructure. We focus on primary outcomes identified from Secondary Activity's logic model to reduce the number of comparisons and minimize the likelihood of false positives. The primary outcomes include student responses on the GRIT scale (Duckworth et al. 2007), regional test scores for 9th grade students and national test scores for 12th grade students, teacher responses on the pedagogical knowledge and attitudes survey, unjustified absences for both students and teachers, and school infrastructure characteristics (condition of the main building, classrooms, and girls' toilets) and management characteristics (management board, school improvement project, and director management training). These outcomes are further defined in Table V.1. We use baseline survey data collected for the impact evaluation as our main data source, except for academic test scores, which come from MASSAR. In Appendix B, Tables B.11–B.19, we present additional findings from exploratory analyses of additional outcomes within each domain.

We present findings separately for lower and upper secondary students, teachers, and schools. Findings from the pooled sample of lower and upper secondary students, teachers, and schools are described in Appendix B, Tables B.7–B.10. Overall, the treatment and control groups are balanced on primary outcomes at baseline, though we find some statistically significant differences between upper secondary students in treatment and control schools and some differences in school infrastructure characteristics at both the lower and upper secondary levels. These differences are generally small in magnitude, however, and will be accounted for by including the characteristics as covariates in our estimation models at endline.

Lower secondary students are equivalent on main outcomes at baseline (Table VI.4). Lower secondary treatment and control students have equivalent grit scores, test scores, and absenteeism rates. Students in both groups scored 2.87 or 2.86 (out of 4) on grit and missed less than a third of a day of school in the past month. Ninth grade students in both groups scored, on average, 8.3 or 8.4 (out of 20) on the regional exam.

Upper secondary level students are also equivalent but have two small differences between treatment and control groups. Treatment students scored slightly lower on grit compared to control students, a difference that is statistically significant at the 5 percent level. Twelfth grade students in treatment schools score slightly lower (on average 0.4 points lower out of 20) on the national exam compared to control students, a difference that is statistically significant at the 1 percent level. Students in both groups missed about half a day of school in the past month.

The baseline level of grit for our sample students is comparable to the levels of grit found for a representative sample of youth in low- and middle-income countries where this soft skill has been measured. The World Bank reports that lower and upper secondary youth in 12 countries scored around 2.5 to 3 on a similar grit scale consisting of 3 items (Roseth et al. 2016). In general, our sample of students also scored better in the five other Big Five Inventory (BFI) soft skills (as reported in Appendix B, Table B.11) than youth in the other countries, although our BFI module included 45 items compared to 16 items in the World Bank surveys. One motivation for the Attahadi model interventions was that soft skills were an important problem to consider for the employability Moroccan youth. Our limited evidence at baseline on self-reported measures of soft skills suggests that Moroccan students in the study sample do not perform more poorly on these skills relative to youth in other developing countries.

Table VI.4. Student primary outcomes, by treatment status

	Mean		Difference (A-B)	p-value (C)
	Treatment (A)	Control (B)		
Lower secondary students				
Grit (range 1 to 4)	2.87	2.86	0.01	0.785
Regional exam (9th grade) (out of 20)	8.3	8.4	-0.0	0.954
Unjustified absences in the past month	0.3	0.3	-0.0	0.754
Sample size lower secondary students (survey)	**840**	**840**	**-**	**-**
Sample size lower secondary schools (survey)	**56**	**56**	**-**	**-**
Sample size lower secondary students (MASSAR)	**14,662**	**68,859**	**-**	**-**
Sample size lower secondary schools (MASSAR)	**56**	**250**	**-**	**-**
Upper secondary students				
Grit (range 1-4)	2.72	2.78	-0.06**	0.037
National exam (12th grade) (out of 20)	10.1	10.5	-0.4***	0.005
Unjustified absences in the past month	0.6	0.5	0.1	0.496
Sample size upper secondary students (survey)	**420**	**420**	**-**	**-**
Sample size upper secondary schools (survey)	**28**	**28**	**-**	**-**
Sample size upper secondary students (MASSAR)	**8,919**	**39,613**	**-**	**-**
Sample size upper secondary schools (MASSAR)	**27**	**133**	**-**	**-**

Source: Grit: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS). Test scores: MASSAR data for the 2017–18 school year. Unjustified absences: school administrative records, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: The grit score reflects the average level of agreement across five statements on a scale of 1 to 4, with 1 being almost always to 4 being almost never. A higher score means that the student exhibits grit to a greater extent. Test scores are out of 20. Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

There is balance on the main teacher outcomes of interest for lower and upper secondary teachers in treatment and control schools (Table VI.5). Teachers at both levels are comparable in terms of their pedagogical knowledge and absenteeism. We assessed teacher pedagogical knowledge and attitudes through questions that asked about the use of 18 different teaching techniques and materials, including discussions, projects, presentations, writing activities, quizzes, interactive lecturing, peer review of assignments, cold calling, technology, and audio-visual aids. On average, both treatment and control lower and upper secondary teachers employ 12 of these 18 techniques and had zero or 0.1 unjustified absences in the past month.

Table VI.5. Main teacher outcomes, by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary teachers				
Score on pedagogical knowledge and attitudes (out of 18)	11.8	12.0	-0.2	0.342
Unjustified absences in the past month	0.0	0.0	0.0	0.827
Sample size lower secondary teachers	**336**	**336**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary teachers				
Score on pedagogical knowledge and attitudes (out of 18)	12.2	12.3	-0.1	0.791
Unjustified absences in the past month	0.0	0.1	-0.1	0.177
Sample size upper secondary teachers	**168**	**168**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Pedagogical knowledge: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS). Unjustified absences: School administrative records, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

The schools in the sample are balanced in terms of the school infrastructure characteristics, with two exceptions related to the quality of the infrastructure (Table VI.6). We observe one small statistically significant difference between treatment and control lower secondary schools and one larger statistically significant difference between upper secondary schools out of six infrastructure characteristics tested.

Lower secondary schools in both groups have low-quality infrastructure. Results from our infrastructure observation show that only 21 percent of treatment schools and 14 percent of

control schools are in excellent or good condition.[16] Treatment and control schools differ on one out of six infrastructure aspects that we measured at baseline. Treatment classrooms have 13.3 out of 16 items on our infrastructure checklist[17] compared to 13.6 for control classrooms, a difference that is statistically significant at the 10 percent level. On average, classrooms in both groups were at a temperature of 19 to 20 degrees Celsius. Girls' toilet facilities in both groups are also similar—the facilities have three out of seven items on our infrastructure checklist, on average.[18]

Upper secondary schools in both groups also score poorly on measures of infrastructure quality. Our infrastructure observation finds that only 25 percent of treatment schools and 11 percent of control schools are in excellent or good condition. Upper secondary schools in both groups differ on one infrastructure characteristic, which is statistically significant at the 5 percent level. All control schools (100 percent), compared to 86 percent of treatment schools, have a complete enclosure wall around the school building. Classrooms in both groups, on average, have 13 or 14 items on our 16-item checklist and were at a temperature of about 20 degrees. Girls toilet facilities in both groups have, on average, three to four out of the seven items on our checklist.

Lower and upper secondary schools have comparable school management characteristics and improvement projects (Table VI.7). There are no statistically significant differences between the two groups on these characteristics, although we might have expected these to differ given that training to improve school management and develop school improvement projects had already been implemented in treatment schools for at least one year by the time of baseline data collection (see Figure II.2 again for the timing of data collection with respect to implementation). Although our study is not well powered to detect impacts at the school level because of the limited number of schools in the evaluation, this result suggests that the Attahadi model interventions related to school management and improvement projects have had minimal impact in schools after one year. One reason for this could be because schools may not have needed assistance on these activities: the data indicate that almost all control schools already have an operational school management board and an improvement project, even with the absence of the Attahadi model. The existing literature also suggests that it can take up to five years before any fundamental changes occur at the school level for school-based management reforms. Because the existence of a school management board and a school improvement project may not capture the full effects of the Attahadi model (indeed, all schools are required by law to have a school

[16] Infrastructure was rated as poor/failed if the building/item cannot continue to perform its original function without significant repairs or is in such a condition that is constitutes a danger for users. A fair rating was assessed if the building/item requires some preventive or curative repairs at a small scale to prevent further deterioration and restore it to its original form. An excellent/good rating means the building/item is sound and has not had any exterior repairs.

[17] The classroom checklist includes the following 16 items: an entrance door, the door is 1 meter wide or greater, the door opens and closes, the door locks, functioning electric lights, at least one window, black or white board, board is visible from the back of the classroom, student tables, student tables are not bolted to the floor, student desks, student desks are not bolted to the floor, chairs for students, chairs are mobile, teacher desk, and platform.

[18] The toilet checklist includes the following items: handicapped accessible, interior doors for all toilet stalls, functioning door, running water, soap, functioning electrical lights, and at least one window.

management board), our implementation study will further investigate whether the quality of school management and the content of school improvement projects have changed as a result of the interventions.

Table VI.6. School infrastructure conditions, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary				
Condition of the main school building				
Overall excellent/good condition (percentage)	21.4	14.3	7.1	0.279
Enclosure for the school building (percentage)	83.9	78.6	5.4	0.442
Rainwater drainage system on roof in excellent/good condition (percentage)	17.3	9.1	8.2	0.137
Condition of the classrooms				
Classroom resources (out of 16)a	13.3	13.6	-0.3*	0.093
Temperature of classroom (in degrees Celsius)	19.4	19.9	-0.5	0.256
Girls toilet facilities				
Toilet facilities conditions (out of 7)b	3.2	3.4	-0.2	0.378
Sample size lower secondary schools	**56**	**56**		
Upper secondary				
Condition of main school building				
Overall excellent/good condition (percentage)	25.0	10.7	14.3	0.143
Enclosure for school building (percentage)	85.7	100.0	-14.3**	0.030
Rainwater drainage system on roof in excellent/good condition (percentage)	21.4	14.3	7.1	0.482
Condition of the classrooms				
Classroom resources (out of 16) a	13.1	13.5	-0.4	0.142
Temperature of classroom (in degrees Celsius)	19.8	19.7	0.1	0.871
Girls toilet facilities				
Toilet facilities conditions (out of 7)b	3.5	3.3	0.2	0.532
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: School infrastructure baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] The classroom checklist includes the following items: entrance door, entrance door is at least one meter wide, entrance door opens and closes, entrance door locks, functioning electric lights, at least once window, black or white board, board can been seen from the back of the classroom, student tables, student tables are not bolted down, student desks, student desks are not bolted down, student chairs, student chairs are mobile, teacher desk, and platform.

[b] The toilet facilities checklist includes the following items: handicapped accessible, interior doors for toilet stalls, functioning door, running water, soap, electric lighting, and at least one window.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Nearly all of the lower secondary schools in both groups have an operational school management board and most (over 85 percent) have a school improvement project (SIP).[19] School management boards in both groups meet at about the same frequency, at least 3-4 times a year. On average, treatment schools allocate 51 percent of their school budget to the SIP and control schools allocate 42 percent. This school budget consists of funds allocated by the state government and coursed through the Regional Academy of Education and Training (AREF), and includes the discretionary funds provided to schools under the project. About half of schools in both groups receive outside sources of funding for the SIP. Common sources of outside funding include Parents of Students Association, Association for School Support, Cultural Sponsorship, and Sports' Association. A minority of school directors in both groups (38 percent of treatment directors and 32 percent of control directors) received management training in the past year.

Almost all of the upper secondary schools in both groups have an operational school management board and a school improvement project. School management boards in the treatment group are 10 percentage points more likely to meet at least 3-4 times in a year than the control group but this difference is not statistically significant. Treatment schools allocate a smaller portion of their school budget to the SIP—42 percent compared to 50 percent for control schools—but this difference is not statistically significant. Over 60 percent of schools in both groups receive outside sources of funding for the SIP. Forty-six percent of treatment directors and 36 percent of control directors received management training in the past year.

[19] Our school survey gathered information on whether the school has a school management board and whether it is operational. Consistent with the government mandate for schools to have a school management board, all school directors reported to have one in the survey, but only a few admitted that they were not operational.

Table VI.7. School management and school improvement projects

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary school				
School management characteristics				
Has an operational school management board (percentage)	100.0	98.2	1.8	0.207
School management board meets at least 3-4 times a year (percentage)	89.2	91.0	-1.8	0.751
School improvement projects (SIP)				
School has a school improvement project (percentage)	87.5	89.3	-1.8	0.771
Percentage of school budget allocated to the project	51.1	42.3	8.8	0.501
School has outside sources of funding for SIP (percentage)	53.6	57.0	-3.4	0.749
Director management training				
Director received management training in the past year (percentage)	38.2	32.2	6.0	0.502
Sample size lower secondary schools	56	56		
Upper secondary school				
School management characteristics				
Has an operational school management board (percentage)	100.0	96.4	3.6	0.312
School management board meets at least 3-4 times a year (percentage)	92.2	81.9	10.4	0.267
School improvement projects (SIP)				
School has a school improvement project (percentage)	100.0	92.9	7.1	0.170
Percentage of school budget allocated to the project	41.7	50.5	-8.8	0.645
School has outside sources of funding for SIP (percentage)	62.1	69.6	-7.5	0.605
Director management training				
Director received management training in the past year (percentage)	46.4	35.7	10.7	0.416
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Director baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

C. Balance tests for treatment and control groups, by subgroups

This section examines balance by student gender for primary student outcomes. Establishing baseline equivalence for this subgroup is important because we plan to examine the endline impacts of the treatment intervention as a function of this characteristic. Baseline equivalence on additional student outcomes by gender are presented in Appendix B, Tables B.20-B.21.

Overall, our results show that the primary outcomes are equivalent between the treatment and control groups for female and male students (Table VI.8). When we detect differences, the magnitudes are relatively small and can be statistically adjusted in our endline analysis.

The primary outcomes of lower secondary male and female students are equivalent across the treatment and control groups. Male and female students in both groups of schools obtained comparable scores on the items related to grit and on the 9th grade regional exam. They also had a similar number of unjustified absences in the past month, according to school records.

There are three statistically significant differences between the treatment and control groups at the upper secondary level: one among male students and two among female students. Upper secondary male students in treatment schools obtained equivalent grit scores and had a similar number of unjustified absences in the past month relative to students in control schools. However, treatment male students correctly answered one half of a question less than male students in the control group—a difference that is statistically significant at the 10 percent level. Female students in treatment and control schools had a similar number of unjustified absences in the past month, but female students in treatment schools obtained lower scores than their counterparts in the control group on items related to grit (a 0.06-point difference) and on the regional exam (a 0.4-percentage point difference). These differences are statistically significant at the 10 and 1 percent levels, respectively.

Table VI.8. Primary student outcomes, by treatment status and gender

	Male						Female					
	Mean			*p*-value	Sample size		Mean			*p*-value	Sample size	
	Treatment (A)	Control (B)	Difference (A-B)	(C)	Treatment (D)	Control (E)	Treatment (F)	Control (G)	Difference (F-G)	(H)	Treatment (I)	Control (J)
Lower secondary students												
Grit score (range 1 to 4)	2.86	2.85	0.01	0.794	444	481	2.89	2.88	0.01	0.798	396	359
Regional exam (9th grade) (out of 20)	8.0	8.0	-0.0	0.850	7,005	33,539	8.7	8.7	0.0	0.973	7,307	33,591
Unjustified absences in the past month	0.4	0.4	0.0	0.763	444	481	0.2	0.2	-0.0	0.456	396	359
Sample size lower secondary students (survey)	**444**	**481**	-	-	-	-	**396**	**359**	-	-	-	-
Sample size lower secondary schools (survey)	**56**	**56**	-	-	-	-	**56**	**56**	-	-	-	-
Sample size lower secondary students (MASSAR)	**7,005**	**33,539**	-	-	-	-	**7,307**	**33,591**	-	-	-	-
Sample size lower secondary schools (MASSAR)	**56**	**249**	-	-	-	-	**56**	**250**	-	-	-	-
Upper secondary students												
Grit score (range 1 to 4)	2.69	2.76	-0.06	0.159	202	204	2.75	2.81	-0.06*	0.094	218	216
National exam (12th grade) (out of 20)	10.0	10.4	-0.4*	0.052	4,127	17,887	10.2	10.6	-0.4***	0.001	4,792	21,726
Unjustified absences in the past month	0.7	0.7	0.0	0.834	202	204	0.4	0.3	0.1	0.359	218	216
Sample size upper secondary students (survey)	**202**	**204**	-	-	-	-	**218**	**216**	-	-	-	-
Sample size upper secondary schools (survey)	**28**	**28**	-	-	-	-	**28**	**28**	-	-	-	-
Sample size upper secondary students (MASSAR)	**4,127**	**17,887**	-	-	-	-	**4,792**	**21,726**	-	-	-	-

	Male						Female					
	Mean		Difference (A-B)	*p*-value (C)	Sample size		Mean		Difference (F-G)	*p*-value (H)	Sample size	
	Treatment (A)	Control (B)			Treatment (D)	Control (E)	Treatment (F)	Control (G)			Treatment (I)	Control (J)
Sample size upper secondary schools (MASSAR)	**28**	**131**	**-**	**-**	**-**	**-**	**28**	**131**	**-**	**-**	**-**	**-**

Source: Grit: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS). Test scores: MASSAR data for the 2017–18 school year. Unjustified absences: school administrative records, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: The grit score reflects the average level of agreement across five statements on a scale of 1 to 4, with 1 being almost always to 4 being almost never. A higher score means that the student exhibits grit to a greater extent. Test scores are out of 20. Columns A and B and F and G present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

This page has been left blank for double-sided copying.

VII. Discussion

The main goals of this report are to document the baseline characteristics of the evaluation sample and assess the degree of similarity in the initial characteristics and outcomes of students, teachers, and schools assigned to the treatment and control groups in the three project regions (TTH, FM, and MS) and examine whether the evaluation sample is representative of the broader population of Moroccan students, teachers, and schools in the TTH region. This chapter summarizes the main findings from the baseline analyses and discusses important threats to the validity of the evaluation, including statistical power to detect intervention impacts and the presence of other school-based programs in the target regions. The chapter also presents the plans for future data collection, analysis, and dissemination of evaluation results.

A. Summary of findings

Overall, we conclude that random assignment worked as intended, creating two groups of students, teachers, and schools that have the same characteristics (on average), except for exposure to the treatment activities. Although results show several statistically significant differences between groups at baseline, the magnitudes of differences are relatively small and can be statistically adjusted for in our analysis. Many of the differences are also likely due to chance, as we performed statistical tests on many characteristics and outcomes. We are confident that random assignment worked as intended because our selection process was conducted through public lotteries in each region that included safeguards to ensure that the selection process was fair, transparent, and free of errors (see Chapter IV for details). The overall balance among the two groups will allow us to estimate the causal impacts of treatment at endline by comparing the outcomes of students and teachers in the two groups and statistically adjusting for any differences we observe at baseline. Moreover, our findings are likely to generalize to the broader population of students, teachers, and schools eligible to receive the Attahadi model in the three study regions.

1. Balance on student characteristics and outcome measures

The evaluation team collected primary survey data from students as well as administrative data from the MASSAR system to examine student characteristics and outcome measures. Our results suggest that the treatment and control groups in both lower secondary and upper secondary schools are balanced on most measures of student characteristics and the main outcomes of interest. We also examined balance between the two groups by student gender on a select number of primary outcomes, including student grit, academic performance, and unjustified absences. The findings suggest that the two groups are comparable.

Although we find balance overall, we observe a few statistically significant differences between the groups on the demographic characteristics and primary outcomes of students. For example, at the lower secondary level, there are slightly more female students in treatment schools than in control schools—a statistically significant difference of four percentage points. We also find that students in treatment schools achieved slightly lower grit scores than control students. Twelfth

grade students in treatment schools also scored slightly lower on the national exam. These differences are small and likely due to chance because placement into each group was based on a randomized design.

2. Balance on teacher characteristics and outcome measures

Our results indicate that lower and upper secondary teachers in treatment and control schools have mostly similar demographic and background characteristics and are also equivalent on all primary outcomes of interest.

However, there are some noteworthy differences between the two groups. Lower secondary treatment schools have a smaller proportion of female teachers than control schools (a difference of seven percentage points) and there are differences on teachers' levels of education that indicate that treatment teachers are more educated, on average, than teachers in the control group. Similarly, upper secondary treatment teachers have three more years of experience than control teachers, are less likely to teach in multiple schools (a difference of five percentage points), are more educated, and are more likely to have a permanent teaching position than teachers in control schools. We will use covariate adjustment to minimize the impact of these differences on our impact estimation models and will take these initial differences into account when interpreting the endline results.

3. Balance on schools and school director characteristics

Schools and school directors in treatment and control schools are equivalent on most characteristics measured at baseline, including the physical condition of the school building, classrooms, and toilet facilities, as well as school directors' gender, length of tenure, and educational background. Treatment lower secondary schools have slightly fewer infrastructure checklist items present in classrooms than control schools, whereas treatment upper secondary schools are less likely to have an enclosure for their school building than control schools. However, we observe no other differences between groups in terms of school infrastructure and school director characteristics. As such, we can be confident that a comparison of outcomes between schools and school directors at endline will reflect impacts of treatment.

4. Representativeness of the evaluation sample

Because we randomly sampled schools, teachers and students for survey data collection among participants in the lotteries in the three study regions, we are confident that the results for our sample at endline will generalize to the broader population of beneficiaries should the Moroccan government decide to expand the interventions to more schools in the three regions. Our findings should also be relevant to schools outside the three regions, given that the three regions were selected to be nationally representative. Using data from the TTH region, we verified that the characteristics of our survey sample of schools, teachers and schools, resemble the broader population of schools, teachers, and schools eligible to receive the Attahadi model in the region.

5. Assessment of evaluation risks

a. Statistical power for the estimation of intervention impacts

In the evaluation design report (Abarcar et al. 2018), we computed minimum detectable effects (MDEs) for the impact evaluation based on our best estimates of sample sizes and other parameters. We revised the MDE calculations using baseline data collected for the evaluation[20] (Table VII.1) and conclude that the evaluation has adequate statistical power to estimate impacts on students and teachers. However, we remain cautious about obtaining precise impact estimates for school-level outcomes. We plan to complement our impact analysis with the qualitative study to help MCC and GoM understand impacts that might occur at the school level.

Under the updated assumptions, the evaluation will be powered to detect impacts as small as 0.16 and 0.19 standard deviations for lower secondary students and teachers, respectively. These are below the effect size of 0.20 that the literature indicates is a reasonable target in the context of education interventions (Damon et al. 2015). We also expect to be able to detect MDEs as small as 0.23 and 0.28 standard deviations for upper secondary student and teacher outcomes, respectively. As we explained in our original design report (Abarcar et al. 2018), these are reasonable effect sizes to aim for given the package of interventions and the large amount of resources provided to each of the selected schools. Our MDEs for school-level impacts remain largely unchanged because the updated intra-class correlations and individual-level sample sizes do not impact those MDEs.

[20] The revised calculations use actual evaluation sample sizes for the Attahadi model students, teachers, and schools, which are slightly smaller at the lower secondary level and slightly larger at the upper secondary level than we had originally assumed. We also use smaller intra-class correlations (ICC) for both student and teacher outcomes (0.07 and 0.04, respectively) based on the evaluation baseline survey data. We continue to assume attrition of 20 percent for students and 10 percent for teachers at endline and the same values for individual- and group-level variance in the outcome explained by covariates for students and teachers.

Table VII.1. Revised MDEs for student-, teacher-, and school-level outcomes, using baseline survey data

	Treatment sample size	Control sample size	MDEs (in standard deviations)		
			Student-level outcomes	Teacher-level outcomes	School-level outcomes
Lower secondary level	56 schools 336 teachers 840 students	56 schools 336 teachers 840 students	0.16 (0.24)	0.19 (0.27)	0.44 (0.44)
Upper secondary level	28 schools 168 teachers 420 students	28 schools 168 teachers 420 students	0.23 (0.35)	0.28 (0.39)	0.63 (0.64)

Note: The original MDEs are shown in parentheses. The revised MDEs assume a two-tailed test with a 95 percent confidence level and 80 percent power. Attrition is assumed at 20 percent for students and 10 percent for teachers at endline. The proportion of individual-level variance in the outcome explained by covariates for students and teachers is assumed to be 0.40 and the proportion of group-level variance explained by covariates is assumed at 0.30. The intra-class correlations (ICC) for students and teachers are 0.07 and 0.04, respectively.

b. Presence of other school-based programs

When education programs in developing countries receive systematically different school-based support from other sources, the support can jeopardize the ability of the impact evaluation to attribute any observable differences to the treatment interventions. The reason for this concern is that the validity of our impact estimates rests on the assumption that random assignment creates equivalent groups, on average, except for implementation of the treatment interventions. If additional interventions are carried out in treatment schools only, then the interpretation of our impact estimates would change to be the impact of the treatment plus additional interventions. However, if the government or other donors provide more support to control schools to compensate for the lack of support from MCC, this would invalidate our estimates of impact because the control group would no longer serve as a valid counterfactual to the treatment group.

We examined the presence of school-based programs in the evaluation schools and find that the two groups of lower and upper secondary schools are statistically equivalent on the support they receive from organizations other than MCC (Table VII.2). However, the differences between the groups are relatively large—12 percentage points in lower secondary schools favoring the control group and 11 percentage points in upper secondary schools favoring the treatment group. These differences may have failed to reach statistical significance due to the evaluation's limited statistical power at the school level. There is also some indication that control group schools at the lower secondary level receive more support from other organizations on infrastructure and extracurricular activities while treatment group schools at the upper secondary level receive slightly fewer outside support than the control group, but more support for other types of school-based programs.

As we describe earlier in this chapter, the use of a randomized evaluation design means that these differences are likely due to chance and are independent from a school's assignment to the

treatment or control condition. However, because baseline data were collected several months after training for the development and use of SIPs and the development of SIPs and partnership fund activities had started, the differences could also reflect early impacts of the intervention. We plan to monitor the implementation of programs other than MCC's throughout the evaluation period to ensure the differences we observe at baseline remain constant. We will also use our qualitative data collection to unpack the type and quantity of support received by schools in the sample.

Table VII.2. Comparison of other school-based programs between treatment and control schools

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary schools				
Receives support from school-based programs other than MCC's (percentage)	55.4	67.9	-12.5	0.148
Number of other organizations providing support	2.1	2.3	-0.2	0.484
Type of school-based program receiving support (percentage)				
Infrastructure	23.2	41.1	-17.9**	0.039
School equipment and books	41.1	33.9	7.1	0.439
Extracurricular activities	30.4	46.4	-16.1*	0.057
Training and management	0.0	3.6	-3.6	0.166
Other	21.4	12.5	8.9	0.213
Sample size lower secondary schools	**56**	**56**		
Upper secondary schools				
Receives support from school-based programs other than MCC's (percentage)	64.3	53.6	10.7	0.431
Number of other organizations providing support	1.4	1.9	-0.5*	0.063
Type of school-based program receiving support (percentage)				
Infrastructure	35.7	42.9	-7.1	0.592
School equipment and books	42.9	42.9	0.0	1.000
Extracurricular activities	46.4	28.6	17.9	0.128
Training and management	3.6	0.0	3.6	0.337
Other	25.0	7.1	17.9*	0.069
Sample size upper secondary schools	**28**	**28**		

Source: School administrative and financial records data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment.

c. Plans for future data collection, analysis, and dissemination

The timing of future data collection activities is yet to be determined given implementation delays. Follow-up quantitative data collection was set to take place in the 2020-2021 school year in TTH and in the 2021–2022 school year in the two other regions (FM and MS), to allow for students to be exposed to program activities for approximately two years (Figure II.2), but given that the COVID-19 pandemic and other factors have delayed program activities, we have decided to delay survey data collection further. This decision was guided by a desire to evaluate impacts on a sample of students who will have more exposure to the full package of interventions. We are actively discussing alternative plans with MCC and will incorporate revisions in an updated Evaluation Design Report once we finalize the timing. We will continue to obtain end-of-year data on students' test scores and progression, repetition, and dropout rates from the MASSAR information system, in addition to our survey data.

Our endline qualitative data collection was set to occur in 2022 in all regions so that findings can be produced and disseminated prior to the compact's closeout and included in GoM discussions regarding scale-up of the program, but we are in discussions to delay this timing, similar to our quantitative surveys[21]. Regardless, we will review documents throughout the course of the evaluation and use the documents to understand the implementation process and timeline as well as any documented challenges to implementation. Staggering the qualitative and quantitative follow-up allows us to provide early results on the process of implementation to MCC, MCA-M, and the MENFPESRS. It also allows us to adapt or focus the quantitative surveys to gather additional data (if needed) to support the impact evaluation.

The timing of the analysis and reporting for the study will be determined by the program's phased rollout schedule (Figure II.2). We expect to present baseline findings discussed in this report in early 2021; the schedule for presenting the qualitative study findings and endline evaluation results will depend on changes in the timing of data collection and evaluation design.

In addition to the final report, we will prepare a policy brief to summarize findings in a concise format, which will make the results more readily accessible and usable to stakeholders and program planners throughout the life of the project. We will work closely with MCC and stakeholders to identify a variety of forums, including conferences, workshops, and publications, to share the results and encourage implementers and policymakers to integrate the findings into future interventions.

[21] The recommendation to collect data in 2022 is driven by the timing of the compact closeout. The literature on dosage—the amount of the intervention required to affect change—shows that there is not a lot of evidence as to how long the interventions should last. The literature is clear that one "touch" or dosage is not enough – and the more consistent the intervention the more likely beneficiaries will uptake the intervention. If the teachers in this program are receiving on-going support and consistent training, then we should be able to see uptake by the endline.

This page has been left blank for double-sided copying.

VIII. Evaluation administration

A. IRB clearance and protocol used for safeguarding human subjects

Mathematica has ensured that the evaluation meets all U.S. research standards for ethical clearance. Mathematica received approval from Health Media Lab, its U.S.-based institutional review board (IRB), before conducting the pretest activity in spring 2018, and provided updated materials for the baseline survey data collection in all three regions before performing those data collection activities. The approval required three sets of documents:

1. A research protocol that described the purpose and design of the research. The document provided information about our plans to protect study participants, their confidentiality and human rights. It also outlined how we acquired consent for their participation in the study.
2. Copies of all data collection instruments and consent forms used for the evaluation.
3. A completed IRB questionnaire that provided information about the research protocol, how we would securely collect and store our data, our plans for protecting participants' rights, and any possible drawbacks for participants that might result from any breach of data confidentiality.

The study qualified for expedited review because it presents minimal risk to participants. The IRB approval was valid for one year and covered the first two rounds of data collection in Tanger-Tétouan-Al Hoceima (May 2018 and April 2019). We submitted a request for renewal before the baseline data collection in Fès-Meknès and Marrakech-Safi in November 2019 and received IRB approval for another year in October 2019. We will continue to seek annual approval until the close of the evaluation.

B. Data quality assurance and data processing

Data quality assurance took place during various stages of the data collection training, fieldwork, and post-fieldwork.

Practice and test at the end of enumerator training. The practice sessions during the enumerator training (in classroom setting and in schools) provided opportunities to build the necessary skill sets before the end of training. Mathematica and DI assessed the enumerators knowledge of the survey instruments, project and respondent specifics, and ability to manage the electronic tablets, as well as their attitude toward other trainees and respondents during practice sessions. Enumerators were also tested at the close of training. We selected the top enumerators based on their performance during the practice and test sessions.

Quality assurance visits during data collection. The enumerators collected survey and observation data using electronic tablets. DI and supervisors ensured compliance with the survey data collection protocols by spot checking the data in the tablets and conducting quality assurance visits to schools during the entire data collection period. Mathematica team members oversaw the first week of each data collection round and worked closely with the supervisors to ensure that they were consistent with providing guidance to enumerators and enforcing study protocols across schools and study regions. The in-field supervision of enumerators focused on

compliance with the survey data collection protocols, such as verifying that enumerators were visiting the correct school and conducting surveys with the correct respondents, following the procedures for replacing students and teachers, administering the consent form in its entirety, reading the survey questions verbatim from their tablets and not changing the content of the questions or judging respondents' answers. Based on our observation, this has resulted in consistent data collection across regions.

Data review during data collection. DI reviewed all survey and school observation data as soon as the data were collected to ensure completeness and consistency. When they found mistakes, they notified the supervisors, who were responsible for re-verifying the data. A few days after the start of data collection, DI began submitting to Mathematica raw survey data daily for preliminary quality assurance review. Data checks included checking that those surveyed were part of the sample, as well as survey completeness, survey logic (for example, whether skip patterns were followed), and data accuracy and consistency. Mathematica then worked directly with DI to resolve any issues encountered to improve data collection during administration of the remaining surveys and observations.

Data processing and quality assurance. Following data collection completion, DI reviewed and cleaned the survey data and submitted the cleaned data files to Mathematica. Mathematica conducted an independent review of the cleaned survey data, which expanded on the data checks conducted during the preliminary review phase. Our internal review included a higher level of detail of the data checks, particularly as it related to skip patterns, data accuracy and consistency. For example, we reviewed all discrete and continuous survey variables for outlying values, coded open text responses into existing response categories (if the response was consistent with one of the survey response options), and expanded the data consistency checks to include additional variables beyond those initially reviewed for consistency. We consulted DI to resolve any data issues we encountered during our review.

C. Data access, privacy, and documentation plan

All data collected for this evaluation were securely transferred from the data collection firm to Mathematica via the platform BOX. BOX is an enterprise cloud-based solution for secure file sharing and collaboration. Each user who participates in the data transfer sets up a unique login credential. Once data are transferred from BOX, Mathematica stores the information on a secure server that is only accessible to the project team who use the data for analysis. After producing and finalizing each of the final evaluation reports, including this baseline report, we will prepare corresponding de-identified data files, user manuals, and codebooks based on the quantitative survey data. We understand that these files could be made available to the public; therefore, the data files, user manuals, and codebooks will be de-identified according to MCC's most recent TREDD guidelines. Public use data files will be free of personal or geographic identifiers that would permit unassisted identification of individual respondents or their households. In addition, we will remove or adjust variables that introduce a reasonable risk of deductive disclosure of the identity of individual participants. We will also recode unique and rare data by using top and bottom coding or by replacing those observations with missing values. If necessary, we will also

collapse any variables that make an individual highly visible, because of geographic or other factors, into less easily identifiable categories.

This page has been left blank for double-sided copying.

References

Abarcar, Paolo, Emilie Bagby, Galina Lapadatova, Caroline Lauver, Audrey Moore, and Matt Sloan. "Evaluation Design Report for the Secondary Education Activity of the Morocco Education & Training Project." Washington, DC: Mathematica Policy Research, July 2018.

Al-Ansari, B. "The Big Five Inventory (BFI): Reliability and Validity of its Arabic Translation in Non Clinical Sample." *European Psychiatry*, no. 33, 2016, pp. S209-S210.

Bagby, Emilie, Anca Dumitrescu, Cara Orfield, and Matt Sloan. "Long-Term Evaluation of the IMAGINE Project in Niger." Washington, DC: Mathematica Policy Research, June 2016.

Bencheikh, Ahmed. "Évaluation de la Première Année de la Mise en Oeuvre des Conventions Cofinancées par le Fonds de Partenariat de l'Éducation pour l'Employabilité Relevant de l'Activité Éducation Scolaire. » Rapport d'évaluation. Volume I. Rabat, Morocco, 2019a.

Bencheikh, Ahmed. "Évaluation de la Première Année de la Mise en Oeuvre des Conventions Cofinancées par le Fonds de Partenariat de l'Éducation pour l'Employabilité Relevant de l'Activité Éducation Scolaire. » Etudes de cas. Volume II. Rabat, Morocco, 2019b.

Blattman, Christopher, and Laura Ralston. "Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs." Available at SSRN, July 2015.

Boller, Kimberly, Stacey Uhl, Lindsay Crozier, Susanne James-Burdumy, Tim Novak, Amy Johnson, and Cheri Vogel. "Using Television as a Teaching Tool: The Impacts of 'Ready to Learn' Workshops on Parents, Educators, and the Children in Their Care." Princeton, NJ: Mathematica Policy Research, 2004.

Bruns, Barbara, Deon Filmer, and Harry A. Patrinos. "Making Schools Work: New Evidence on Accountability Reforms." Washington, DC: World Bank, February 2011.

Burnett, C. "The 'Uptake' of a Sport-for-Development Programme in South Africa." *Sport, Education & Society,* vol. 20, no. 7, 2015, pp. 819-837.

Cid, A. "Interventions using Regular Activities to Engage High-Risk School-Age Youth: A Review of After-School Programs in Latin America and the Caribbean." *Prevention Science,* vol. 18, no. 7, 2017, pp. 879-886.

Conn, Katherine. "Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Rigorous Impact Evaluations." Unpublished doctoral thesis. New York, NY: Columbia University, 2014.

Cuesta, Ana, Paul Glewwe, and Brooke Krause. "School Infrastructure and Educational Outcomes: A Literature Review, with Special Reference to Latin America." *Economia*, vol. 17, no. 1, fall 2016, pp. 95–130.

Damon, Amy, Paul Glewwe, Suzanne Wisniewski, and Bixuan Sun. "What Education Policies and Programs Affect Learning and Time in School in Developing Countries? A Review of Academic and Aid Agency Evaluations from 1990 to 2014 for the Swedish Expert Group for Aid Studies (EBA)." Stockholm: Expert Group for Aid Studies, December 2015.

Duckworth, Angela, Christopher Peterson, Michael Matthews, and Dennis Kelly. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology*, vol. 92, 2007, pp. 1087–101. 10.1037/0022-3514.92.6.1087.

Duflo, Esther, Pascalina Dupas, and Michael Kremer. "Peer Effects, Pupil-Teacher Ratios, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya." September 14, 2007. Available at https://pdfs.semanticscholar.org/6e56/71179d88ba1810706760a275c2759e3d890e.pdf.

Durlak, J.A., R.P. Weissberg, A.B. Dymnicki, R.D. Taylor, and K.B. Schellinger. "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-analysis of School-Based Universal Interventions." *Child Development,* vol. 82, no. 1, 2011, pp. 405–432.

Evans, David K., and Anna Popova. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systemic Reviews." World Bank Policy Research Working Paper No. 7203. Hyattsville, MD: World Bank Group, February 2015.

Gaible, E. "After School Programme for Girls and Girls Access to Education Programme." Kathmandu, Nepal: UNICEF Nepal, 2010.

Gates, Sarah, Laura Lippman, Noel Shadowen, Holly Burke, Obed Diener, and Morrisa Malkin. "Key Soft Skills for Cross-Sectoral Youth Outcomes." Washington, DC: United States Agency for International Development, YouthPower: Implementation, YouthPower Action, November 2016.

Gertler, P., H.A. Patrinos, and E. Rodríguez-Oreggia. "Parental Empowerment in Mexico: Randomized Experiment of the Apoyo a la Gestión Escolar (AGE) in Rural Primary Schools in Mexico: Preliminary Findings." Unpublished manuscript. World Bank, Washington, DC, 2010.

Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath. "The Impact of Soft Skills Training on Female Youth Employment: Evidence from a Randomized Experiment in Jordan." *IZA Journal of Labor and Development,* vol. 5, no. 9, 2016.

Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." In *Education Policy in Developing Countries*, edited by Paul Glewwe. Chicago and London: University of Chicago Press, 2014.

Ibarraran, Pablo, Laura Ripani, Bibiana Taboada, Juan Miguel Villa, and Brigida Garcia. "Life Skills, Employability, and Training for Disadvantaged Youth: Evidence from a Randomized Evaluation Design." *IZA Journal of Labor and Development,* vol. 3, no. 10, 2014, pp. 10.

Kautz, Tim, James J. Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success." National Bureau of Economic Research. No. 20749. December 2014.

Khan, M. Y. and A. Jamil. "Perceptual Posture of Stakeholders: Efficacy of Sports in Curbing Violent Behaviors among Youth." *Journal of Educational Research (1027-9776),* vol. 20, no. 2, 2017, pp. 170-183.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. "The Challenge of Education and Learning in the Developing World." *Science* (New York, N.Y.), vol. 340, 2013, pp. 297–300. 10.1126/science.1235350.

Krishnaratne, Shari, Howard White, and Ella Carpenter. "Quality Education for All Children? What Works in Education in Developing Countries." New Delhi: International Initiative for Impact Evaluation (3ie), Working Paper 20, 2013.

Land, Tony. "Implementing Institutional and Capacity Development: Conceptual and Operational Issues." Discussion Paper No. 14. Maastricht, Netherlands: European Centre for Development Policy Management, March 2000.

Levy, Dan, Matt Sloan, Leigh Linden, and Harounan Kazianga. "Impact Evaluation of Burkina Faso's BRIGHT Program." Washington, DC: Mathematica Policy Research, June 2009.

Maebuta, J. "The Education Pacific Islands Children Deserve: The Learn and Play Project in the Solomon Islands." *International Education Journal: Comparative Perspectives,* vol. 10, no. 2, 2011, pp. 99-112.

McEwan, Patrick. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research,* vol. 85, 2015, pp. 353–394. 10.3102/0034654314553127.

Millennium Challenge Corporation (MCC). Compact with the Government of Morocco, 2015a. https://assets.mcc.gov/content/uploads/2017/05/compact-morocco-employability-and-land.pdf

Millennium Challenge Corporation (MCC). "Morocco's Growth Diagnostic. Identifying Morocco's Binding Constraints to Broad-Based Growth." Washington, DC: Millennium Challenge Corporation, October 2015b.

Muller Mariano, F. A. and da Silva Filho, L. S. P. "Craque do Amanhã: Sports, Schools, and Family." *Gestión y Política Pública,* vol. 24, 2015, pp. 139-169.

Mullis, I.V.S., M.O. Martin, P. Foy, and M. Hooper. "TIMSS 2015 International Results in Mathematics." 2016a. Available at Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Mullis, I.V.S., M.O. Martin, P. Foy, and M. Hooper. "TIMSS 2015 International Results in Science." 2016b. Available at Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/.

Murnane, Richard J., and Alejandro J. Ganimian. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations." Unpublished manuscript, 2014.

OECD. "PISA 2018 Results (Volume I): What Students Know and Can Do." Paris: OECD Publishing, 2019. Available at https://doi.org/10.1787/5f07c754-en.

Plaisant, O., R. Courtois, C. Réveillère, G.A. Mendelsohn, and O.P. John. "Validation par Analyse Factorielle du Big Five Inventory Français (BFI-Fr). Analyse Convergente avec le NEO-PI-R." *Annales Médico-Psychologiques*, no. 168, 2010, pp. 97–106.

Popova, A., D.K. Evans, M.E. Breeding, and V. Arancibia. "Teacher Professional Development Around the World: The Gap Between Evidence and Practice." World Bank, 2018.

Raikes, Helen H., Julia C. Torquati, Susan Hegland, H. Abigail Raikes, Jacqueline Scott, Lana Messner, Carla Peterson, Kathy Thornburg, Becky Houf, and Sandra Scott. "Studying the Culture of Quality Early Education and Care: A Cumulative Approach to Measuring Characteristics of the Workforce and Relations to Quality in Four Midwestern States." In *Critical Issues in Early Childhood Professional Development*, edited by M. Zaslow and I. Martinez-Beck. Baltimore, MD: Brookes, 2006.

Rankin, Kristen, Drew B. Cameron, Kara Ingraham, Anjini Mishra, Jane Burke, Mario Picon, Jorge Miranda, and Annette N. Brown. "Youth and Transferable Skills: An Evidence Gap Map." Evidence Gap Map Report 2. New Delhi: International Initiative for Impact Evaluation (3ie), 2015.

Roseth, Viviana V., Alexandria Valerio, and Marcela Gutierrez. "Education, Skills, and Labor Market Outcomes: Results from Large-Scale Adult Skills Surveys in Urban Areas in 12 Countries." Report. Washington, DC: World Bank Group, 2016.

Seow, Poh-Sun, and Gary Pan. "A Literature Review of the Impact of Extracurricular Activities Participation on Students' Academic Performance." *Journal of Education for Business* vol. 89, no. 7, 2014, pp. 361-366.

Solutions for Youth Employment (S4YE). "Toward Solutions for Youth Employment: A 2015 Baseline Report." Washington DC: Solutions for Youth Employment (S4YE), 2015.

TradingEconomics. "Morocco Youth Unemployment Rate." 2019. Available at tradingeconomics.com/morocco/youth-unemployment-rate?continent=asia.

UNESCO Institute of Statistics. "Data for the Sustainable Development Goals: Morocco." Available at http://uis.unesco.org/country/MA#slideoutmenu. Accessed December 16, 2019.

United Nations. "World Youth Report: Youth and the 2030 Agenda for Sustainable Development," 2018.

USAID. "Enhancing Youth Employability." Available at https://2017-2020.usaid.gov/morocco/fact-sheets/employability-youth-enhanced. Accessed December 16, 2019.

Yamauchi, Futoshi. "An Alternative Estimate of School-Based Management Impacts on Students' Achievements: Evidence from the Philippines." Policy Research Working Paper No. WPS 6747, Impact Evaluation Series No. IE 113. Washington, DC: World Bank Group, 2014.

Winton, Pamela J., and Jeanette A. McCollum. "Preparing and Supporting High Quality Early Childhood Practitioners: Issues and Evidence." In *Preparing and Supporting Effective Practitioners: Evidence and Applications in Early Childhood and Early Intervention*, edited by P. J. Winton, J. A. McCollum, and C. Catlett. Washington, DC: Zero to Three Press, 2008

World Bank. "World Development Report 2018: Learning to Realize Education's Promise." Washington, DC: World Bank, 2018.

This page has been left blank for double-sided copying.

Appendix A

Stakeholder comments

This page has been left blank for double-sided copying.

Memo

To: Ryan Moore and Carolyn Perrin

From: Emilie Bagby, Matt Sloan, and Galina Lapadatova

Date: September 20, 2022

Subject: Baseline Report revisions: Independent Evaluation Services in Support of the Secondary Activity of the Morocco Compact's Education and Training for Employability Project

---

Thank you to the Secondary Activity of the Morocco Compact's Education and Training for Employability Project stakeholders for taking the time to make the thoughtful written comments and suggestions on the Baseline Report. This memo shares our responses to these written comments. We organize our responses by question/comment. Please contact us if you have any questions or need more information. We also welcome any verbal questions and discussion about the Baseline Report or other study documents. Page numbers below refer to pages in the Baseline Report with tracked changes.

I. Comments from MCA-Morocco M&E

- **Comment on page 3:** The Evaluation of the Student Achievement and Education Management Information System (EMIS/MASSAR). Other interventions include: improving ongoing monitoring and remediation practices; improving initial teacher training in assessment; developing a school performance measurement framework

  Mathematica response: Thank you for your comment. We've added the three interventions to the table II.1 on page 3.

- **Comment on page 12:** The sub-component "assessment of student learning" is not mentioned, is it not covered by the implementation study?

  Mathematica response: Yes, the sub-component will be evaluated in the implementation study at the system level. The specific research question (number 11) is noted in table IV.1 on page 12: *How have the interventions improved student assessment, data and policy feedback in the MASSAR system, leading to a more performance-oriented education system?*

II. Comments from M&E from TTH region

- **Comment on page 7:** In reviewing the document, the link between school performance and the school autonomous management was explained by determinants that were not foreseen under the "Secondary education" project. Management autonomy as described on page 8 of the report far exceeds that of the schools receiving the project. In our case, autonomy was translated or manifested simply by providing schools with leeway in the financial management of the PEI. Thus, the reasoning discussed in the report will only make sense if management autonomy within the framework of the project extends to cover other aspects of management, including administrative and HR management.

Mathematica response: We agree with this statement, and we have noted that autonomous School Management reforms vary according to the level of autonomy granted to the school and the degree of involvement of other parties concerned (such as local authorities, parents and other members of the community), which makes it difficult to draw general conclusions.

- **Comment on page 7:** Stakeholder engagement must also be reasoned internally. In fact, if the members of the Schools' *Comité de pilotage* (CP) show resistance to the project or abandon their involvement in the implementation or monitoring and evaluation of the PEI, this would have an impact on the degree of achievement of the results. That said, the non-engagement of internal or external stakeholders with the schools constitutes a risk that will compromise the purpose of the project.

  Mathematica response: We agree, and we plan to ask a question about difficulties the director may have encountered in deploying and implementing PEI activities. We also plan to discuss these issues qualitatively during our interview and focus group data collection.

- **Comment on page 11:** The report considers the student, the teacher, the school and the system as being the units of measurement for the evaluation of the sub-activities "Attahadi Model". It also seems important to add the school director to this list and to add questions for the director related to evaluation and design given the essential role played by the director in achieving the results of the project.

  Mathematica response: We agree that the school director is an important actor in the school system. Our questionnaires include one for the school director, and as there is 1 per school, we report the responses from that questionnaire with the other school-level outcomes.

- **Comment on page 17:** Among the sources of the basic data, there was also an administrative and financial data booklet which was administered to the school director to collect information.

  Mathematica response: Yes, this is correct. The administrative and financial data booklet was part of the director survey instrument.

- **Comment on page 18:** Isn't it relevant to draw on the reference framework adopted by the Ministry of National Education (MEN) for the definition of transverse skills since we work in a Moroccan context?

  Mathematica response: With an RCT design the baseline measures do not need to be the same as endline, but they do need to be correlated. We will work with the local actors to identify the appropriate endline measures of soft skills; we added a footnote to page 18.

- **Comment page 39:** Replace PEI by PE since the design of the PEI (designed and conducted according to the DEPART process) came with the project, whereas before the start of the project and even during the baseline data collection, there was just the notion of the PE (designed and conducted according to the EPAR process).

  Mathematica response: The concept of the PEI was already established at the time of baseline data collection and the PEI training had already initiated when we conducted our data collection.

- **Comment on page 39:** Replace "school the management committee" with "the school management board" and the "plan" with "project" to be in line with the nomenclature adopted by the Moroccan education system.

  Mathematica response: Thank you for the comment. We have made the suggested changes throughout the baseline report.

This page has been left blank for double-sided copying.

Appendix B

Supplementary tables

This page has been left blank for double-sided copying.

While the main analysis of primary characteristics and outcomes presented in Chapter VI shows relative balance between the treatment and control groups, in this appendix we present supplemental analysis to show the equivalence of the two groups across the full spectrum of indicators we collected. Given the sheer number of indicators presented, we would expect to see imbalance across many due to chance. For this reason, we focused on a smaller set of core indicators for the primary analysis presented in Chapter VI. Where we do detect statistically significant differences between the treatment and control groups, we will include the relevant variables in our endline estimation models to correct for any imbalances.

A. Demographic and academic characteristics

In this section, we present additional analyses of the evaluation sample and assess whether the demographic and academic characteristics of students, teachers, and schools are equivalent between treatment and control groups at baseline (Tables B.1–B.6). We present the tables with the balance tests for the findings discussed in Chapter VI, as well as analysis of additional demographic and academic characteristics. For students, these additional characteristics include the language spoken at home, parent's educational attainment, home assets, and travel to school. For teachers, we present supplemental balance tests on their use of quizzes and assessment, what activities increase instruction time, reasons for absences, and professional development training. At the school level, we additionally examine language of instruction, school closures, and support from businesses and organizations for renovations.

Table B.1 presents the balance tests for the student demographic and academic characteristics discussed in Chapter VI Section B.

Table B.1. Primary student demographic and academic characteristics, by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary students				
Age (years)	13.0	13.0	0.0	0.891
Female (percentage)	47.1	42.7	4.4**	0.046
Language spoken at home (percentage)				
Arabic	90.1	89.6	0.5	0.868
French	0.4	1.0	-0.6*	0.084
Tamazight (Berber)	9.4	9.3	0.1	0.966
Father's highest educational attainment (percentage)				
None	29.0	29.9	-0.9	0.691
Koranic school	10.7	10.7	-0.1	0.962
Primary school	26.6	25.3	1.3	0.555
Lower secondary school	13.7	14.8	-1.1	0.497
Upper secondary school	11.5	13.4	-1.9	0.337
Postsecondary school	7.2	4.5	2.7*	0.066

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Other	1.3	1.4	-0.2	0.732
Mother's highest educational attainment (percentage)				
None	48.0	51.3	-3.3	0.247
Koranic school	1.4	1.2	0.3	0.631
Primary school	23.3	20.6	2.7	0.176
Lower secondary school	11.2	13.4	-2.2	0.147
Upper secondary school	9.6	8.8	0.8	0.668
Post-secondary school	4.1	3.5	0.6	0.657
Other	2.3	1.2	1.2	0.115
Sample size lower secondary students	**840**	**840**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary students				
Age (years)	15.9	15.9	0.1	0.490
Female (percentage)	51.9	51.4	0.5	0.891
Language spoken at home (percentage)				
Arabic	95.7	91.9	3.8	0.152
French	0.2	0.7	-0.5	0.282
Tamazight (Berber)	3.6	6.9	-3.3	0.194
Father's highest educational attainment (percentage)				
None	19.1	23.4	-4.3	0.239
Koranic school	9.7	10.6	-0.9	0.674
Primary school	23.6	28.9	-5.3	0.133
Lower secondary school	15.7	138.	1.9	0.493
Upper secondary school	18.5	14.2	4.3	0.136
Postsecondary school	11.4	7.1	4.3	0.114
Other	2.1	2.1	0.0	0.981
Mother's highest educational attainment (percentage)				
None	41.0	49.4	-8.4*	0.092
Koranic school	0.8	1.2	-0.5	0.486
Primary school	20.5	19.9	0.7	0.793
Lower secondary school	15.5	11.7	3.9*	0.076
Upper secondary school	11.1	11.4	-0.3	0.907
Postsecondary school	8.3	5.4	2.9	0.215
Other	2.8	1.0	1.8**	0.049
10th grade track (percentage)				
Original	1.0	1.0	0.0	1.000
General	94.5	97.4	-2.9	0.152
Technical	0.5	0.2	0.2	0.533
Professional	4.0	1.4	2.6	0.159
Sample size upper secondary students	**420**	**420**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data. Koranic is considered a preschool and not part of the official public education system.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.2 presents findings on additional student baseline characteristics. Among lower secondary students of both groups there is one statistically significant difference, at the one percent level, related to the average amount of time it takes students to travel to school. For upper secondary students, we observe four statistically significant differences, at the five and 10 percent levels, between the treatment and control groups related to home assets (computer and internet access) and mode of travel to school (school transportation and car). These differences are small in magnitude and can be adjusted for in our endline estimation models.

Table B.2. Additional student characteristics, by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary students				
Home assets (percentage)				
Radio	47.5	47.4	0.1	0.963
Television	98.8	98.6	0.2	0.639
Computer	32.7	33.3	-0.6	0.817
Internet connection	28.3	28.8	-0.5	0.842
Telephone landline	24.6	24.6	0.0	1.000
Running water	93.6	93.5	0.1	0.942
Refrigerator	96.9	97.5	-0.6	0.420
Number of rooms at home	6.1	6.1	-0.1	0.648
Has mobile phone (percentage)	38.0	41.8	-3.8	0.150
Has books at home (aside from schoolbooks) (percentage)	60.1	62.3	-2.1	0.465
Minutes it takes for students to get to the school	26.1	22.3	3.8***	0.007
Mode of travel to school (percentage)				
Foot	75.5	76.0	-0.5	0.848
Public transport	5.0	4.4	0.6	0.630
Bike	4.3	4.0	0.2	0.879
School transportation	17.1	15.6	1.5	0.383
Car (including taxi)	4.9	4.9	0.0	1.000
Other	0.4	-0.0	0.4	0.278
Student reports that the school offers transportation (percentage)	38.2	34.2	4.0	0.213
Student feels safe on the way to school (percentage)	68.2	70.8	-2.6	0.297
Sample size lower secondary students	**840**	**840**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary students				
Home assets (percentage)				
Radio	51.9	51.0	1.0	0.819
Television	99.0	98.8	0.2	0.687
Computer	53.3	44.8	8.6*	0.053
Internet connection	48.1	40.0	8.1**	0.048
Telephone landline	35.0	29.0	6.0	0.168
Running water	95.0	91.7	3.3	0.230
Refrigerator	97.9	99.0	-1.2	0.130
Number of rooms at home	6.5	6.8	-0.3	0.178
Has mobile phone (percentage)	82.6	78.1	4.5	0.155
Has books at home (aside from schoolbooks)	74.8	76.2	-1.4	0.720
Minutes it takes for students to get to the school	24.6	25.0	-0.4	0.853

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Mode of travel to school (percentage)				
Foot	79.0	76.4	2.6	0.578
Public transport	7.4	9.5	-2.1	0.444
Bike	5.2	4.0	1.2	0.454
School transportation	6.4	13.6	-7.1*	0.051
Car (including taxi)	8.1	3.8	4.3**	0.048
Other	0.7	-0.0	0.7	0.125
Student reports that the school offers transportation (percentage)	22.6	34.0	-11.4	0.102
Student feels safe on the way to school (percentage)	72.9	70.9	2.0	0.498
Sample size upper secondary students	**420**	**420**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.3 presents the balance tests for teachers' demographic and employment characteristics discussed in Chapter VI Section B.

Table B.3. Primary teacher demographic and employment characteristics, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary teachers				
Female (percentage)	48.5	55.7	-7.1*	0.072
Years of teaching in the school	6.9	6.9	0.1	0.925
Years of teaching in total	15.3	15.5	-0.3	0.734
Teaches in different schools (percentage)	5.4	8.0	-2.7	0.161
Highest level of education (percentage)				
Upper secondary	-0.0	0.9	-0.9*	0.065
First university cycle (DEUG)[a]	5.7	6.3	-0.6	0.747
Second university cycle (*licence*)	56.3	63.1	-6.8*	0.091
Teacher's certification	21.1	19.0	2.1	0.559
Third university cycle (master's)	16.4	8.9	7.4***	0.001
Doctorate	0.6	1.8	-1.2	0.115
Teacher position (percentage)				
Permanent	81.0	82.1	-1.2	0.704
Contract	18.8	17.9	0.9	0.775
Other	0.6	0.0	0.6	0.123

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Sample size lower secondary teachers	**336**	**336**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary teachers				
Female (percentage)	33.9	38.1	-4.2	0.329
Years of teaching in the school	7.8	6.3	1.5	0.117
Years of teaching in total	17.2	14.1	3.0**	0.032
Teaches in different schools (percentage)	5.4	10.7	-5.3*	0.076
Highest level of education (percentage)				
Upper secondary	0.0	0.0	0.0	1.000
First university cycle (DEUG)[a]	1.8	-0.0	1.8**	0.036
Second university cycle (*licence*)	41.1	48.8	-7.7*	0.056
Teacher's certification	13.7	10.7	3.0	0.377
Third university cycle (master's)	39.9	35.7	7.4	0.340
Doctorate	3.6	4.8	-1.2	0.594
Teacher position (percentage)				
Permanent	94.5	85.8	8.7**	0.019
Contract	5.5	14.2	-8.7**	0.019
Other	0.6	0.0	0.6	0.260
Teacher track (percentage)				
General	100	100	1.5	1.000
Technical	1.8	3.0	-1.2	0.548
Professional	4.8	3.0	1.8	0.594
Original	0.0	0.0	0.0	1.000
Sample size upper secondary teachers	**168**	**168**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] DEUG = Diplôme d'études universitaires générales.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.4 presents additional teacher characteristics related to the use of student tests, quizzes, and assessments, reasons for absences from school, and professional development training. Consistent with what we observed for primary teacher demographic and employment characteristics, we find a number of statistically significant differences between teachers in the two groups. Among lower secondary teachers, there are five differences, statistically significant at the five and one percent levels. These differences are related to how teachers use tests or quizzes, if they use assessments to compare students' performance to other students in Morocco, and their participation in professional development. Among upper secondary teachers there are three differences, statistically significant at the five or 10 percent levels, for teachers use of tests

or quizzes, use of tests or quizzes to plan lessons, and their participation in professional development.

Table B.4. Additional teacher characteristics, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary teachers				
Uses tests or quizzes in the classroom (percentage)	96.4	96.1	0.3	0.843
Uses information from tests or quizzes to (percentage)				
Help understand what students are learning	76.9	76.1	0.9	0.840
Help plan lessons	45.1	56.3	-11.2***	0.007
Help provide grades	38.6	52.0	-13.3***	0.002
Help understand individual weaknesses	80.8	89.2	-8.4**	0.012
Uses information from assessments to (percentage)				
Help understand what students are learning	42.0	43.1	-1.1	0.942
Help compare students' performance to other students in neighboring schools	59.4	71.8	-12.4	0.594
Help compare students' performance to other students in Morocco	60.8	25.6	35.2**	0.027
Help understand individual weaknesses	36.4	38.8	-2.4	0.879
Other	3.4	1.4	2.0	0.533
Main reasons for absence in school (percentage)				
Family obligations	18.6	17.4	1.2	0.898
School closures	14.4	15.5	-1.0	0.874
Illness	56.1	56.8	-0.7	0.939
Other	12.2	11.1	1.1	0.864
Participated in professional development in the school year (percentage)	25.1	13.7	11.3***	0.001
Sample size lower secondary teachers	**336**	**336**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary teachers				
Uses tests or quizzes in the classroom (percentage)	97.0	92.9	4.2*	0.060
Uses information from tests or quizzes to				
Help understand what students are learning	63.8	65.4	-1.6	0.788
Help plan lessons	48.1	35.6	12.4*	0.080
Help provide grades	50.9	44.3	6.7	0.380
Help understand individual weaknesses	85.3	80.1	5.3	0.405
Uses information from assessments to				
Help understand what students are learning	38.0	53.8	-15.8	0.442
Help compare students' performance to other students in neighboring schools	55.4	46.7	8.7	0.679
Help compare students' performance to other students in Morocco	27.1	51.6	-24.5	0.255

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Help understand individual weaknesses	20.8	24.3	-3.6	0.747
Other	0.0	0.0	0.0	1.000
Main reasons for absence in school (percentage)				
Family obligations	25.2	16.6	8.6	0.563
School closures	20.6	15.9	4.6	0.240
Illness	54.5	56.3	-1.7	0.918
Other	5.1	11.4	-6.3	0.547
Participated in professional development in the school year (percentage)	22.3	13.1	9.2**	0.047
Sample size upper secondary teachers	**168**	**168**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.5 presents the balance tests for school and director characteristics discussed in Chapter VI Section B.

Table B.5. Primary school and director characteristics, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary school				
Number of students enrolled	873	842	30	0.651
Age of school (years)	25.4	22.2	3.3	0.356
Located in rural area (percentage)	23.2	30.4	-7.1	0.186
Offers boarding to students (percentage)	25.0	10.7	14.3**	0.010
Number of students the school can board	129.6	134.6	-5.0	0.876
Annual budget (excluding AREF) for 2017-2018 (in MAD)	28,632	33,243	-4,611	0.651
School director				
Female (percentage)	3.6	7.1	-3.6	0.430
Length of tenure as director (years)	8.7	7.8	0.9	0.264
Teaching experience (years)	14.3	15.0	-0.7	0.570
Highest level of education (percentage)				
Upper secondary	0.0	1.8	-1.8	0.303
First university cycle (DEUG)	3.6	1.8	1.8	0.553
Second university cycle (*licence*)	76.8	73.2	3.6	0.654
Teacher's certification	14.3	12.5	1.8	0.752
Third university cycle (master's)	5.4	10.7	-5.4	0.312

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Doctorate	0.0	0.0	0.0	1.000
Other	0.0	0.0	0.0	1.000
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary school				
Number of students enrolled	1,041	1,006	35	0.761
Age of school (years)	22.7	25.4	-2.7	0.663
Located in rural area (percentage)	32.1	35.7	-3.6	0.682
Offers boarding to students (percentage)	10.7	21.4	-10.7	0.184
Number of students the school can board	368.0	209.5	158.5	0.695
Annual budget (excluding AREF) for 2017–2018 (in MAD)	28,539	24,494	4,045	0.732
School director				
Female (percentage)	3.6	-0.0	3.6	0.349
Length of tenure as director (years)	6.4	6.3	0.1	0.879
Teaching experience (years)	14.0	15.8	-1.9	0.313
Highest level of education (percentage)				
Upper secondary	3.6	7.1	-3.6	0.583
First university cycle (DEUG)	0.0	0.0	0.0	1.000
Second university cycle (*licence*)	71.4	57.1	14.3	0.291
Teacher's certification	10.7	21.4	10.7	0.254
Third university cycle (master's)	14.3	7.1	7.1	0.401
Doctorate	-0.0	3.6	-3.6	0.312
Other	-0.0	-3.6	-3.6	0.312
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Director baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: MAD = Moroccan dirhams. Columns A and B present regression-adjusted means that account for the strata used in random assignment.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.6 presents findings for additional school characteristics. There are no statistically significant differences between lower secondary treatment and control schools. At the upper secondary level, there are two statistically significant differences—for the percentage of schools that use French as a main language in the classroom and the number of days the school was closed. Both differences are significant at the 10 percent level only.

Table B.6. Additional school characteristics, by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary school				
Main language(s) used in the classroom (percentage)				
Arabic	100.0	100.0	0.0	1.00
French	57.1	66.1	-8.9	0.197
Tamazight (Berber)	0.0	1.8	-1.8	0.329
Days school was closed in the academic year (excluding holidays and weekends)	0.1	0.1	-0.0	0.879
Receives support from businesses or local organizations for building renovations or new equipment (percentage)	7.1	14.3	-7.1	0.234
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary school				
Main language(s) used in the classroom (percentage)				
Arabic	100.0	100.0	0.0	1.000
French	82.1	64.3	17.9*	0.054
Tamazight (Berber)	0.0	0.0	0.0	1.000
Days school was closed in the academic year (excluding holidays and weekends)	-0.0	0.3	-0.3*	0.085
Receives support from businesses or local organizations for building renovation or new equipment (percentage)	7.1	3.6	3.6	0.580
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Director baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS). Administrative and financial survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

B. Pooled analysis

Although the primary interest of the evaluation is in estimating impacts at each of the LS and US school levels, our evaluation will also explore pooled impacts, combining the data on students, teachers, and schools on both levels. The advantage of this approach is that it improves on statistical power by utilizing all available data to estimate impacts. A pooled analysis will enable us to detect small but meaningful impacts that might not appear to be statistically significant when we analyze lower and upper secondary schools separately. In this section, we show baseline equivalence on primary outcomes for students, teachers, and schools for the pooled samples (Tables B.7–B.10).

Table B.7 present the pooled analysis for primary student outcomes. There are no statistically significant differences between treatment and control students. We are, however, unable to provide pooled analysis for student test scores because they are grade-level specific.

Table B.7. Student primary outcomes by treatment status, for the full sample of upper and lower secondary students

	Mean		Difference (A-B)	p-value (C)
	Treatment (A)	Control (B)		
Lower and upper secondary students				
Grit (range 1 to 4)	2.82	2.84	-0.01	0.426
Unjustified absences in the past month	0.4	0.4	0.0	0.838
Sample size students	**1,260**	**1,260**	**-**	**-**
Sample size schools	**84**	**84**	**-**	**-**

Source: Grit: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS). Unjustified absences: school administrative records data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: For grit, the score reflects the average level of agreement (on a scale of 1 to 4, with 1 being almost always to 4 being almost never) across 5 statements. The minimum score is 1 and the maximum score is 4. A higher score means that the student exhibits that soft skill or grit to a greater extent. Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

Table B.8 present the pooled analysis for primary teacher outcomes. There are no statistically significant differences between treatment and control teachers.

Table B.8. Teacher primary outcomes by treatment status, for the full sample of upper and lower secondary teachers

	Mean		Difference (A-B)	p-value (C)
	Treatment (A)	Control (B)		
Lower and upper secondary teachers				
Score on pedagogical knowledge and attitudes (out of 18)	11.9	12.1	-0.1	0.355
Unjustified absences in the past month	0.0	0.1	-0.0	0.252
Sample size teachers	**504**	**504**	**-**	**-**
Sample size schools	**84**	**84**	**-**	**-**

Source: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

Table B.9 present the pooled analysis for school infrastructure characteristics. There are two statistically significant differences between treatment and control schools—the percentage of schools that are in excellent or good condition (significant at the 10 percent level) and the average number of classroom resources (significant at the five percent level).

Table B.9. School infrastructure characteristics by treatment status, for the full sample of upper and lower secondary schools

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower and upper secondary school				
Condition of the main school building				
Overall excellent/good condition (percentage)	22.6	13.1	9.5*	0.080
Enclosure for the school building (percentage)	84.5	85.7	-1.2	0.817
Rainwater drainage system on roof in excellent/good condition (percentage)	18.8	10.9	7.8	0.119
Condition of the classrooms				
Classroom resources (out of 16)[a]	13.2	13.6	-0.4**	0.025
Temperature of classrooms (in degrees Celsius)	19.5	19.8	-0.3	0.388
Girls toilet facilities				
Toilet facilities conditions (out of 7)[b]	3.3	3.4	-0.1	0.728
Sample size schools	**84**	**84**	**-**	**-**

Source: School infrastructure baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] The classroom checklist includes the following items: entrance door, entrance door is at least 1 meter wide, entrance door opens and closes, entrance door locks, functioning electric lights, at least once window, black or white board, board can been seen from the back of the classroom, student tables, student tables are not bolted down, student desks, student desks are not bolted down, student chairs, student chairs are mobile, teacher desk and platform.

[b] The toilet facilities checklist includes the following items: handicapped-accessible, interior doors for toilet stalls, functioning door, running water, soap, electric lighting, at least one window.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.10 present the pooled analysis for school management and improvement projects. There are no statistically significant differences between treatment and control schools.

Table B.10. School management and improvement projects by treatment status, for the full sample of upper and lower secondary schools

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower and upper secondary school				
School management characteristics				
Has an operational school management board (percentage)	100.0	97.6	2.4	0.110
School management board meets at least 3-4 times a year (percentage)	90.2	88.0	2.2	0.657
School improvement projects (SIP)				
School has a school improvement project (percentage)	91.7	90.5	1.2	0.789
Percentage of school budget allocated to the project	48.7	44.8	3.9	0.713

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
School has outside sources of funding for SIP (percentage)	56.4	61.3	-4.8	0.575
Director management training				
Director received management training in the past year (percentage)	40.9	33.3	7.6	0.303
Sample size schools	**84**	**84**	**-**	**-**

Source: Director baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

C. Additional outcomes of interest

In this section, we assess balance between the treatment and control groups for secondary outcomes that can impact long-term student outcomes (Tables B.11–B.19). For students, we consider additional soft skills, test scores, measures of absenteeism, progression, graduation, repetition, and dropout rates, academic aspirations, and gender equality perceptions. For teachers, we present findings on additional measures of absenteeism, time spent on different academic activities, and gender equality perceptions. Finally, we present analysis of additional school infrastructure characteristics and school budget allocation and spending priorities.

Table B.11 presents findings for additional student soft skills, test scores, and measures of absenteeism. There are three statistically significant differences between lower secondary students in both groups for soft skills. Differences for openness and extroversion are statistically significant at the five percent level and the difference for conscientiousness is statistically significant at the one percent level, but are small in magnitude. At the upper secondary level, there are two statistically significant differences—for neuroticism and the 12th grade passing grade—between treatment and control students. These differences are significant at the one and five percent levels, respectively, but are small in magnitude.

Table B.11. Additional soft skills, test scores, and absenteeism measures, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary students				
BFI soft skills (range 1 to 5)				
Openness	3.68	3.63	0.05**	0.036
Conscientiousness	3.82	3.75	0.07***	0.003
Extroversion	3.48	3.44	0.04**	0.029
Agreeableness	3.96	3.94	0.02	0.284
Neuroticism	2.52	2.54	-0.02	0.294

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Test scores (out of 20)				
Local exam (9th grade)	11.0	11.1	-0.1	0.480
Passing grade (*moyenne passage*) (9th grade)	10.3	10.4	-0.0	0.663
Absences				
Days absent in the school in the last two weeks (self-report)	0.6	0.6	-0.1	0.317
Justified absences in the past month (days)	0.1	0.1	-0.0	0.496
Sample size lower secondary students (survey)	**840**	**840**	**-**	**-**
Sample size lower secondary schools (survey)	**56**	**56**	**-**	**-**
Sample size lower secondary students (MASSAR)	**14,662**	**68,859**	**-**	**-**
Sample size lower secondary schools (MASSAR)	**56**	**250**	**-**	**-**
Upper secondary students				
BFI soft skills (range 1 to 5)				
Openness	3.59	3.57	0.02	0.498
Conscientiousness	3.55	3.59	-0.04	0.148
Extroversion	3.43	3.41	0.02	0.582
Agreeableness	3.92	3.92	-0.00	0.918
Neuroticism	2.74	2.65	0.09***	0.008
Test scores (out of 20)				
Passing grade (*moyenne passage*) (12th grade)	10.8	11.1	-0.3**	0.036
Absences				
Days absent in the school in the last two weeks (self-report)	0.9	0.8	0.1	0.563
Justified absences in the past month (days)	0.2	0.1	0.0	0.885
Sample size upper secondary students (survey)	**420**	**420**	**-**	**-**
Sample size upper secondary schools (survey)	**28**	**28**	**-**	**-**
Sample size upper secondary students (MASSAR)	**8,919**	**39,613**	**-**	**-**
Sample size upper secondary schools (MASSAR)	**27**	**133**	**-**	**-**

Source: BFI soft skills and self-reported absences: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS). Test scores: MASSAR data for the 2017–18 school year. Justified absences: school administrative records, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: For each BFI (Big Five Inventory) soft skill, scores reflect the average level of agreement (on a scale of 1 to 5, with 1 being strongly disagree and 5 being strongly agree) across 8–10 statements related to that soft skill. The minimum score is 1 and the maximum score is 5. A higher score means that the student exhibits that personality trait or grit to a greater extent. Test scores are out of 20. Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/** Difference between treatment and control group means is statistically significant at the .01/.05 level.

Table B.12 provides balance tests for students' progression, graduation, repetition and dropout rates, by grade level. Among lower secondary students, we observe statistically significant differences, at the five percent level, between the treatment and control groups for 7th grade progression and repetition rates. Among upper secondary students, we find six statistically significant differences between the treatment and control groups for 10th and 11th grade

progression, repetition, and dropout rates. These differences are significant at the five and 10 percent levels. While we observe several statistically significant difference, they are all relatively small in magnitude.

Table B.12. Student progression, graduation, repetition, and dropout rates, by treatment status

	Mean		Difference (A-B)	*p*-value (C)	Sample size	
	Treatment (A)	Control (B)			Treatment (D)	Control (E)
Lower secondary students						
Progression and graduation (percentage)						
7th grade students progressed to 8th grade for the 2018–19 school year	64.5	67.4	-2.8**	0.041	17,854	76,278
8th grade students progressed to 9th grade for the 2018–19 school year	70.7	72.3	-1.6	0.222	13,858	62,919
9th grade students progressed to 10th grade for the 2018–19 year	52.9	54.5	-1.6	0.387	14,662	68,859
9th grade students graduating this school year	60.0	62.1	-2.1	0.449	54	53
Repetition (percentage)						
7th grade students repeating the grade for the 2018–19 school year	22.6	20.6	2.0**	0.037	17,854	76,278
8th grade students repeating the grade for the 2018–19 school year	17.8	17.2	0.6	0.541	13,858	62,919
9th grade students repeating the grade for the 2018–19 school year	29.7	29.6	0.1	0.925	14,662	68,859
Dropout (percentage)						
7th grade students dropping out during or between school years	15.2	14.5	0.7	0.306	17,854	76,278
8th grade students dropping out during or between school years	13.4	12.4	1.0	0.140	13,858	62,919
9th grade students dropping out during or between school years	18.9	17.5	1.4	0.187	14,662	68,859
Upper secondary students						
Progression and graduation (percentage)						
10th grade students progressed to 11th grade for the 2018–19 school year	86.5	83.8	2.7**	0.026	8,914	43,918
11th grade students progressed to 12th grade for the 2018–19 school year	85.4	83.2	2.2*	0.052	8,677	41,914
12th grade students graduating this school year	69.5	73.2	-3.7	0.349	26	25
Repetition (percentage)						
10th grade students repeating the grade for the 2018–19 school year	7.5	9.1	-1.6*	0.064	8,914	43,918
11th grade students repeating the grade for the 2018–19 school year	8.5	9.9	-1.4*	0.094	8,677	41,914
Dropout (percentage)						
10th grade students dropping out during or between school years	7.1	8.5	-1.5**	0.025	8,914	43,918
11th grade students dropping out during or between school years	7.0	8.1	-1.1*	0.071	8,677	41,914

Source: Progression, repetition, and dropout rates: MASSAR data for the 2017–18 and 2018–19 school years. Graduation rates for 9th and 12th grade students: School administrative records data from May 2018 in TTH and November/December 2019 in FM and MS.

Notes: Progression and repetition rates capture whether or not the student progressed to the next grade level or repeated the same grade level for the 2018–19 school year. Dropout rates include students who dropped

out during or between the 2017–18 and 2018–19 school years. Graduation rates capture whether or not the student graduated 9th or 12th grade at the end of the 2017–18 school year in TTH and at the end of the 2018–19 school year in FM and MS. in TTH, these data were collected in May 2018 and provide rates for the 2017–18 school year. Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

**/* Difference between treatment and control group means is statistically significant at the .05/.10 level.

Table B.13 presents students' academic aspirations—plans to enroll in upper secondary school for lower secondary students and plans to complete upper secondary school and continue to university for upper secondary students. Lower secondary students in the treatment and control groups are equivalent. For upper secondary students, there is one statistically significant difference between the two groups. The percentage of students who plan to study a university discipline other than science, technology, engineering and math (STEM), law, economics, administration, social science, literature, Islamic studies, or languages, is lower in the treatment group.[22] This difference is significant at the 5 percent level and small in magnitude.

[22] Among students who indicate that they plan to study a different discipline at university, chosen fields of study include police academy, vocational training, and physical education.

Table B.13. Student academic aspirations, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary students				
Plans to enroll in upper secondary school (percentage)	98.1	97.5	0.6	0.384
Track in which student plans to enroll in upper secondary school (percentage)				
General	89.2	91.4	-2.2	0.268
Technical	4.1	3.7	0.4	0.766
Professional	5.2	4.1	1.1	0.407
Original	1.5	0.7	0.7	0.249
Sample size lower secondary students	**840**	**840**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary students				
Plans to complete upper secondary school (percentage)	98.0	97.7	0.3	0.751
Plans to enroll in university (percentage)	86.9	88.6	-1.8	0.502
Discipline student plans to study in university (percentage)				
STEM field	47.5	45.9	1.6	0.700
Law/economics/administration/social sciences	41.1	34.3	6.9	0.108
Literature	6.0	8.8	-2.8	0.238
Islamic studies	1.2	2.3	-1.1	0.419
Languages	2.9	3.6	-0.7	0.612
Other	1.2	5.1	-3.9**	0.012
Sample size upper secondary students	**420**	**420**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.14 presents findings for student perceptions on 19 items related to gender equality. There are four statistically significant differences among lower secondary students in both groups. These differences are for preference for female teachers and agreement with the following statements: boys and girls have the same opportunity to take math and science classes, girls and boys interact equally in classroom discussions, and women have the same right as men to become teachers. These differences are significant at the five and 10 percent levels and are relatively small in magnitude. Among upper secondary students, there is only one statistically significant difference, at the ten percent level, between treatment and control groups. Thus, we conclude that the two groups of upper secondary students are equivalent on these measures of gender equality perceptions.

Table B.14. Gender equality perceptions among students, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary students				
Student's teachers are: (percentage)				
Primarily men	32.8	29.0	3.8	0.190
Primarily women	53.1	56.1	-3.0	0.320
Both men and women	14.1	14.9	-0.8	0.661
Preferred gender for teachers (percentage)				
Male	26.0	23.7	2.3	0.305
Female	32.1	36.7	-4.5*	0.078
No preference	41.9	39.6	2.3	0.326
Agrees with the following statements: (percentage)				
Girls have the same right to go to school as boys	94.3	93.9	0.4	0.771
Girls and boys have the same right to enroll in higher levels of education	96.8	96.7	0.1	0.876
Girls and boys are encouraged to choose subjects they are interested in taking in school	91.0	91.3	-0.4	0.784
Girls can succeed in math and science	91.8	91.4	0.4	0.775
Girls can have careers in math and science	91.1	91.7	-0.6	0.611
Boys and girls have the same opportunity to take math and science classes	91.9	94.4	-2.5**	0.043
Our teachers encourage girls and boys to take math and science classes	91.9	92.3	-0.4	0.783
My parents encourage me to look at careers in math and science fields	89.4	90.1	-0.7	0.632
Girls and boys interact in discussions equally in my classrooms	70.1	74.2	-4.0*	0.097
Teachers at my school encourage boys and girls to participate in class discussions	91.0	89.6	1.3	0.371
Women have the same right as men to become teachers	97.7	99.0	-1.3**	0.021
A female community leader can be as effective as a male community leader	89.8	89.9	-0.1	0.935
Women have the right to hold leadership positions in the community.	91.8	93.5	-1.7	0.200
Sample size lower secondary students	**840**	**840**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary students				
Student's teachers are: (percentage)				
Primarily men	34.1	35.8	-1.7	0.590
Primarily women	51.0	44.8	6.2	0.178
Both men and women	14.9	19.4	-4.5	0.108
Preferred gender for teachers (percentage)				
Male	37.4	36.4	1.0	0.771
Female	21.4	24.5	-3.1	0.231

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
No preference	41.2	39.0	2.1	0.532
Agrees with the following statements: (percentage)				
Girls have the same right to go to school as boys	93.3	95.2	-1.9	0.195
Girls and boys have the same right to enroll in higher levels of education	98.8	97.4	1.4	0.122
Girls and boys are encouraged to choose subjects they are interested in taking in school	85.5	89.5	-4.0*	0.057
Girls can succeed in math and science	96.2	96.7	-0.5	0.653
Girls can have careers in math and science	96.9	95.0	1.9	0.109
Boys and girls have the same opportunity to take math and science classes	95.0	95.0	0.0	1.000
Our teachers encourage girls and boys to take math and science classes	81.7	83.6	-1.9	0.544
My parents encourage me to look at careers in math and science fields	75.5	75.5	0.0	1.000
Girls and boys interact in discussions equally in my classrooms	63.3	67.4	-4.0	0.264
Teachers at my school encourage boys and girls to participate in class discussions	87.1	89.8	-2.6	0.257
Women have the same right as men to become teachers	98.3	98.8	-0.5	0.509
A female community leader can be as effective as a male community leader	91.4	92.4	-1.0	0.604
Women have the right to hold leadership positions in the community	94.8	94.3	0.5	0.720
Sample size upper secondary students	**420**	**420**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.15 presents balance tests for additional teacher outcomes including time use and absences. Among lower secondary teachers in both groups, we observe one statistically significant difference, at the five percent level, for time spent planning and preparing lessons. For upper secondary teachers, there are three statistically significant differences between treatment and control teachers with respect to their time use—specifically the amount of time spent meeting with the school director or other teachers, meeting with parents, and planning tests or quizzes. These differences are significant at the five and 10 percent levels and are small in magnitude.

Table B.15. Additional teacher outcomes, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary teachers				
Time spent in school per day (hours)	4.1	4.2	-0.1	0.202
Days absent from school in the last two weeks (self-report)	0.3	0.3	0.0	1.000
Justified absences in the past month	0.5	0.5	0.0	0.902
Total hours devoted to all school-related activities at home or in school per week[a]	59.5	60.4	-1.0	0.654
Time devoted to school-related activities at home or in school per week (hours)				
Planning and preparing lessons	6.3	5.4	0.9**	0.036
Teaching students	20.8	20.1	0.7	0.167
Grading students' tests and homework	4.6	5.2	-0.6	0.273
Meeting with director or other teachers	1.0	0.9	0.1	0.288
Meeting with parents of students	0.6	0.4	0.1	0.431
Planning tests or quizzes for students	3.7	4.1	-0.5	0.127
Administering tests or quizzes to students	4.1	4.3	-0.2	0.490
Correcting students' tests or quizzes	6.6	7.3	-0.7	0.209
Preparing students for national assessments outside of school curriculum	1.4	1.8	-0.5	0.102
Providing additional academic support for underperforming students	1.4	1.6	-0.3	0.169
Supervising extracurricular activities	1.1	1.1	0.0	0.947
Sample size lower secondary teachers	**336**	**336**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary teachers				
Time spent in school per day (hours)	4.1	4.2	-0.0	0.412
Days absent from school in the last two weeks (self-report)	0.2	0.3	-0.1	0.462
Justified absences in the past month	0.3	0.3	-0.0	0.758
Total hours devoted to all school-related activities at home or in school per week[a]	65.0	65.5	-0.6	0.905
Time devoted to school-related activities at home or in school per week (hours)				
Planning and preparing lessons	6.8	7.5	-0.7	0.231
Teaching students	18.0	17.8	0.1	0.781
Grading students' tests and homework	4.8	6.4	-1.5	0.143
Meeting with director or other teachers	1.2	0.9	0.3**	0.034
Meeting with parents of students	0.5	0.2	0.3*	0.069
Planning tests or quizzes for students	5.2	4.3	0.9*	0.096
Administering tests or quizzes to students	4.8	4.2	0.6	0.274
Correcting students' tests or quizzes	7.2	8.0	-0.8	0.318
Preparing students for national assessments outside of school curriculum	1.9	2.1	-0.2	0.650

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Providing additional academic support for underperforming students	1.9	1.6	0.3	0.231
Supervising extracurricular activities	1.1	1.0	0.1	0.594
Sample size upper secondary teachers	**168**	**168**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] School-related activities include planning and preparing lessons, teaching students, grading students' tests or homework, meeting with director or other teachers, meeting with parents of students, planning tests or quizzes for students, correcting students' tests or quizzes, preparing students for national assessments outside of school curriculum, providing additional academic support for underperforming students, and supervising extracurricular activities.

Table B. 16 presents equivalence findings for teachers' gender equality perceptions, as measured by their agreement with 13 different statements. We find that lower secondary teachers in treatment and control schools are equivalent, as there is only one statistically significant difference, at the 10 percent level, between both groups. In upper secondary schools, there are six statistically significant differences between teachers in the two groups. One of these differences, for agreement with the statement that girls can succeed in math and science, is significant at the one percent level. The other five differences (agreement that girls can have careers in math and science, boys and girls have the same opportunity to take math and science classes, girls and boys interact equally in class discussions, teachers encourage girls and boys to participate in class, and a female community leader can be as effective as a male leader) are statistically significant at the five and 10 percent levels. Most of these differences are small in magnitude.

Table B.16. Gender equality perceptions among teachers, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary teachers				
Agrees with the following statements: (percentage)				
Girls have the same right to go to school as boys	92.9	94.3	-1.5	0.478
Girls and boys have the same right to enroll in higher levels of education	94.6	94.9	-0.3	0.867
Girls and boys are encouraged to choose subjects they are interested in taking in school	88.1	92.3	-4.2*	0.056
Girls can succeed in math and science	98.8	98.8	0.0	1.000
Girls can have careers in math and science	100.0	99.7	0.3	0.258
Boys and girls have the same opportunity to take math and science classes	96.7	98.5	-1.8	0.127

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Our teachers encourage girls and boys to take math and science classes	97.6	98.5	-0.9	0.317
Girls and boys interact in discussions equally in my classrooms	74.7	76.5	-1.8	0.580
Teachers at my school encourage boys and girls to participate in class discussions	97.6	98.2	-0.6	0.594
A female community leader can be as effective as a male community leader	91.4	92.3	-0.9	0.648
Women have the same right as men to become teachers	99.1	99.4	-0.3	0.630
Students at my school prefer female teachers	33.9	36.9	-3.0	0.351
Women have the right to hold leadership positions in the community	95.8	97.0	-1.2	0.337
Sample size lower secondary teachers	**336**	**336**	**-**	**-**
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary teachers				
Agrees with the following statements: (percentage)				
Girls have the same right to go to school as boys	95.2	91.7	3.6	0.202
Girls and boys have the same right to enroll in higher levels of education	98.8	96.4	2.4	0.145
Girls and boys are encouraged to choose subjects they are interested in taking in school	91.7	90.5	1.2	0.716
Girls can succeed in math and science	100.0	96.4	3.6***	0.002
Girls can have careers in math and science	100.0	97.6	2.4**	0.024
Boys and girls have the same opportunity to take math and science classes	98.8	95.2	3.6**	0.043
Our teachers encourage girls and boys to take math and science classes	96.4	97.0	-0.6	0.750
Girls and boys interact in discussions equally in my classrooms	79.8	67.9	11.9**	0.010
Teachers at my school encourage boys and girls to participate in class discussions	98.2	94.6	3.6*	0.079
A female community leader can be as effective as a male community leader	95.8	91.1	4.8**	0.027
Women have the same right as men to become teachers	98.2	98.2	0.0	1.000
Students at my school prefer female teachers	31.5	28.6	3.0	0.544
Women have the right to hold leadership positions in the community	96.4	95.8	0.6	0.770
Sample size upper secondary teachers	**168**	**168**	**-**	**-**
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Teacher baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/*: Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.17 presents balance tests for additional school infrastructure characteristics. There are three statistically significant differences between lower secondary treatment and control schools.

These differences—separate spaces for faculty/admin personnel, separate toilet facilities for boys and girls, and number of toilet stalls—are significant at the five and ten percent levels and are relatively small in magnitude. At the upper secondary level, we observe two statistically significant differences, at the ten percent level only. These differences are for the average number of toilet stalls available and the percentage of schools in each group with toilet facilities inside the school premises.

Table B.17. Additional school infrastructure characteristics, by treatment status

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Lower secondary				
Condition of the main school building				
Wheelchair access ramps (percentage)	25.0	30.4	-5.4	0.493
Separate space for faculty/admin personnel outside of classroom (percentage)	94.6	85.7	8.9*	0.082
Main entrance doors in excellent/good condition (percentage)	48.2	36.9	11.3	0.288
Exterior walls of building painted (percentage)	98.2	94.6	3.6	0.290
Interior walls of building painted (percentage)	97.9	94.9	3.0	0.535
Condition of the classrooms				
Students sit at individual desks (percentage)	1.8	3.6	-1.8	0.511
Condition of WASH facilities				
Separate toilet facilities for boys and girls (percentage)	100.0	94.6	5.4**	0.041
Flush toilet sewer connection (percentage)	51.8	62.5	-10.7	0.151
Girls toilet facilities				
Toilet facilities inside school premises (percentage)	46.5	47.1	-0.6	0.945
Number of toilet stalls available for students	6.1	5.6	0.5	0.230
Boys toilet facilities				
Toilet facilities conditions (out of 7)a	3.1	3.1	-0.0	0.887
Toilet facilities inside school premises (percentage)	35.5	36.1	-0.6	0.940
Number of toilet stalls available for students	7.3	5.7	1.6*	0.095
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary				
Condition of the main school building				
Wheelchair access ramps (percentage)	32.1	39.3	-7.1	0.597
Separate space for faculty/admin personnel outside of classroom (percentage)	100.0	96.4	3.6	0.214
Main entrance doors in excellent/good condition (percentage)	36.1	33.6	2.5	0.848
Exterior walls of building painted (percentage)	92.9	96.4	-3.6	0.537
Interior walls of building painted (percentage)	99.6	89.4	10.3	0.125
Condition of the classrooms				
Students sit at individual desks (percentage)	7.1	10.7	-3.6	0.312

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Condition of WASH facilities				
Separate toilet facilities for boys and girls (percentage)	100.0	100.0	0.0	1.000
Flush toilet sewer connection (percentage)	64.3	67.9	-3.6	0.779
Girls toilet facilities				
Toilet facilities inside school premises (percentage)	32.1	39.3	-7.1	0.505
Number of toilet stalls available for students	6.6	7.8	-1.3*	0.070
Boys toilet facilities				
Toilet facilities conditions (out of 7)a	3.1	3.1	0.0	0.922
Toilet facilities inside school premises (percentage)	17.9	35.7	-17.9*	0.064
Number of toilet stalls available for students	5.8	6.4	-0.5	0.560
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: School infrastructure baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: The toilet facilities checklist includes the following items: handicapped-accessible, interior doors for toilet stalls, functioning door, running water, soap, electric lighting, at least one window. Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

[a] The toilet facilities checklist includes the following items: handicapped-accessible, interior doors for toilet stalls, functioning door, running water, soap, electric lighting, at least one window.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.18 presents findings for additional school infrastructure characteristics—specifically related to the science lab, computer room, library, and recreation area. Among lower secondary schools in both groups, there is only one statistically significant difference, at the five percent level, for the percentage of libraries with books. There are no statistically significant differences between upper secondary treatment and control schools.

Table B.18. Additional school infrastructure characteristics (science lab, computer room, library, and recreation area), by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary schools				
Science classroom				
School has a science classroom (percentage)	100.0	96.4	3.6	0.166
Score on checklist of 9 items in the classroom	7.3	7.4	-0.1	0.734
Science rooms with equipment in excellent/good condition (percentage)	5.5	5.5	-0.0	1.000
Computer room				
School has a computer room (percentage)	62.5	57.1	5.4	0.520
Number of computers in the computer room	12.9	11.8	1.1	0.465
Computer rooms with equipment in excellent/good condition (percentage)	17.3	28.9	-11.6	0.323

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Library, resource room, or multimedia room				
School has a school library, resource room, or multimedia room (percentage)	71.4	71.4	0.0	1.000
Room has books (percentage)	71.6	92.1	-20.4**	0.020
Books in new/good condition (percentage)	15.0	25.6	-10.6	0.180
Outdoor recreation area				
School has an outdoor space for sports (percentage)	98.2	100.0	-1.8	0.207
Recreational equipment is in excellent/good condition (percentage)	3.6	9.0	-5.5	0.266
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary schools				
Science classroom				
School has a science classroom (percentage)	96.4	100.0	-3.6	0.337
Score on checklist of 9 items in the classroom	7.4	7.5	-0.1	0.802
Science rooms with equipment in excellent/good condition (percentage)	-0.1	7.2	-7.3	0.132
Computer room				
School has a computer room (percentage)	71.4	78.6	-7.1	0.503
Number of computers in the computer room	13.2	14.6	-1.4	0.538
Computer rooms with equipment in excellent/good condition (percentage)	13.2	16.8	-3.7	0.778
Library, resource room, or multimedia room				
School has a school library, resource room, or multimedia room (percentage)	50.0	67.9	-17.9	0.114
Room has books (percentage)	80.7	96.9	-16.2	0.172
Books in new/good condition (percentage)	6.9	21.6	-14.7	0.415
Outdoor recreation area				
Schools has an outdoor space for sports (percentage)	100.0	96.4	3.6	0.312
Recreational equipment is in excellent/good condition (percentage)	3.7	7.3	-3.7	0.557
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: School infrastructure baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: Columns A and B present regression-adjusted means that account for the strata used in random assignment. Sample sizes shown are for the largest sample, but some regressions may include a smaller sample size due to missing data.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.19 presents balance tests for schools' allocation of the AREF budget and top three spending priorities. At the lower secondary level, there are six statistically significant differences between schools in both groups, out of 28 characteristics tested. Two of these differences, percentage of AREF budget spent on extracurricular clubs and percentage of schools with extracurricular clubs as a top spending priority, are significant at the five percent level. The other four differences are significant at the 10 percent level. At the upper secondary level, there are four statistically significant differences, all of which are significant at the ten percent level.

Table B.19. School budget allocation and spending priorities, by treatment status

	Mean		Difference (A-B)	*p*-value (C)
	Treatment (A)	Control (B)		
Lower secondary school				
Percentage of AREF budget spent on:				
School materials	2.8	4.2	-1.4	0.111
Overhead	1.0	0.5	0.4	0.408
Building maintenance	1.8	3.0	-1.2	0.620
Emergency maintenance	0.8	0.1	0.8	0.109
Building renovations	2.0	2.7	-0.7	0.776
Management and partnership	0.3	0.5	-0.1	0.705
Equipment	3.1	1.3	1.8*	0.099
Social support	12.8	9.1	3.7*	0.094
Academic support	0.4	0.6	-0.2	0.651
Extracurricular clubs	2.8	0.5	2.3**	0.033
Professional orientation	2.1	0.1	2.0	0.207
Public services (electricity, water, sanitation)	26.8	26.7	0.1	0.985
Student meals	10.9	14.8	-3.9	0.194
Gardening and cleaning	15.2	18.1	-3.0	0.292
Other	17.6	17.9	-0.3	0.864
Top three spending priorities of the school include (percentage):				
School materials	8.9	19.6	-10.7*	0.091
Overhead	30.4	41.1	-10.7	0.228
School building maintenance	21.4	28.6	-7.1	0.374
Emergency maintenance	19.6	12.5	7.1	0.314
School building renovations	8.9	10.7	-1.8	0.741
Management and partnership	0.0	1.8	-1.8	0.352
Equipment	48.2	55.4	-7.1	0.473
Social support	21.4	8.9	12.5*	0.062
Student support services	42.9	42.9	0.0	1.000
Extracurricular clubs	64.3	44.6	19.6**	0.034
Professional orientation	5.4	5.4	0.0	1.000
Public services	14.3	8.9	5.4	0.381
Student meals	1.8	3.6	-1.8	0.501
Sample size lower secondary schools	**56**	**56**	**-**	**-**
Upper secondary school				
Percentage of AREF budget spent on:				
School materials	1.0	1.7	-0.8	0.444
Overhead	0.9	0.3	0.5	0.379
Building maintenance	9.1	0.1	9.0*	0.059
Emergency maintenance	0.4	0.7	-0.4	0.596
Building renovations	2.7	2.6	0.1	0.974

	Mean			
	Treatment (A)	Control (B)	Difference (A-B)	*p*-value (C)
Management and partnership	0.1	0.1	0.0	1.000
Equipment	3.0	3.8	-0.9	0.569
Social support	3.6	8.0	-4.4*	0.082
Academic support	4.1	0.1	3.9	0.158
Extracurricular clubs	3.7	0.9	2.8*	0.086
Professional orientation	2.6	0.1	2.5	0.324
Public services (electricity, water, sanitation)	24.5	31.4	-6.9	0.226
Student meals	8.2	13.5	-5.3	0.411
Gardening and cleaning	22.4	21.9	0.5	0.874
Other	14.0	14.7	-0.7	0.882
Top three spending priorities of the school include (percentage):				
School materials	14.3	25.0	-10.7	0.327
Overhead	28.6	17.9	10.7	0.303
School building maintenance	17.9	35.7	-17.9	0.134
Emergency maintenance	21.4	14.3	7.1	0.503
School building renovations	-0.0	7.1	-7.1	0.114
Management and partnership	7.1	7.1	0.0	1.000
Equipment	46.4	64.3	-17.9	0.146
Social support	21.4	25.0	-3.6	0.739
Student support services	46.4	32.1	14.3	0.271
Extracurricular clubs	60.7	39.3	21.4*	0.092
Professional orientation	7.1	7.1	0.0	1.000
Public services	14.3	10.7	3.6	0.652
Student meals	0.0	0.0	0.0	1.000
Sample size upper secondary schools	**28**	**28**	**-**	**-**

Source: Director baseline survey data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: AREF = Regional Academy of Education and Training (*Académie Régionale d'Education et de Formation).* Columns A and B present regression-adjusted means that account for the strata used in random assignment.

**/* Difference between treatment and control group means is statistically significant at the .05/.10 level.

D. Additional subgroup analysis

This section presents findings by student gender for supplementary student outcomes—additional soft skills, test scores, and measures of absenteeism, as well as progression, repetition, and dropout rates (Tables B.20–B.21). Establishing baseline equivalence for this subgroup is important because we plan to examine the endline impacts of the treatment intervention as a function of this characteristic.

Table B.20 presents findings by gender for additional student soft skills, test scores, and absenteeism measures. There are no statistically significant difference between male students in both groups in lower secondary schools. We observe four statistically significant differences between female students in treatment and control schools for soft skills. One of these differences,

for conscientiousness, is statistically significant at the one percent level. The other three, openness, extroversion, and agreeableness, are significant at the five percent level. At the upper secondary level, there are two statistically significant differences between male students in treatment and control schools—conscientiousness (significant at the five percent level) and the 12th grade passing grade (significant at the 10 percent level). There are also two statistically significant differences between upper secondary female students in both groups—neuroticism (significant at the one percent level) and 12th grade passing grade (significant at the 10 percent level). While these differences are all small in magnitude.

Table B.20. Additional student soft skills, test scores, and measures of absenteeism, by treatment status and gender

	Male						Female					
	Mean				Sample size		Mean				Sample size	
	Treatment (A)	Control (B)	Difference (A-B)	p-value (C)	Treatment (D)	Control (E)	Treatment (F)	Control (G)	Difference (F-G)	p-value (H)	Treatment (I)	Control (J)
Lower secondary students												
BFI soft skills (range 1 to 5)												
Openness	3.62	3.60	0.02	0.539	444	481	3.74	3.66	0.07**	0.017	396	359
Conscientiousness	3.77	3.74	0.03	0.270	444	481	3.89	3.77	0.11***	0.000	396	359
Extroversion	3.46	3.45	0.01	0.507	444	481	3.50	3.43	0.07**	0.041	396	359
Agreeableness	3.89	3.91	-0.02	0.466	444	481	4.05	3.97	0.08**	0.015	396	359
Neuroticism	2.46	2.49	-0.04	0.249	444	481	2.58	2.59	-0.01	0.700	396	359
Test scores (out of 20)												
Local exam (9th grade)	10.7	10.8	-0.1	0.544	7,185	34,449	11.3	11.4	-0.1	0.431	7,477	34,410
Passing grade (moyenne passage) (9th grade)	9.9	10.0	-0.1	0.582	7,005	33,539	10.7	10.7	-0.0	0.732	7,307	33,591
Absences												
Days absent from school in the last two weeks (self-report)	0.7	0.7	-0.1	0.275	444	481	0.5	0.5	0.0	0.731	396	359
Justified absences from school in the past month	0.1	0.1	-0.0	0.341	444	481	0.1	0.1	0.0	0.728	396	359
Upper secondary students												
BFI soft skills (range 1 to 5)												
Openness	3.55	3.57	-0.02	0.614	202	204	3.64	3.57	0.07	0.109	218	216
Conscientiousness	3.50	3.59	-0.09**	0.037	202	204	3.60	3.60	-0.00	0.960	218	216
Extroversion	3.45	3.47	-0.02	0.598	202	204	3.40	3.37	0.04	0.368	218	216
Agreeableness	3.87	3.88	-0.01	0.818	202	204	3.96	3.97	-0.01	0.845	218	216
Neuroticism	2.60	2.53	0.07	0.118	202	204	2.89	2.75	0.14***	0.007	218	216

	Male						Female					
	Mean				Sample size		Mean				Sample size	
	Treatment (A)	Control (B)	Difference (A-B)	p-value (C)	Treatment (D)	Control (E)	Treatment (F)	Control (G)	Difference (F-G)	p-value (H)	Treatment (I)	Control (J)
Test scores (out of 20)												
Passing grade (*moyenne passage*) (12th grade)	10.5	10.8	-0.3*	0.056	4,127	17,887	11.1	11.3	-0.3*	0.058	4,792	21,726
Absences												
Days absent from school in the last two weeks (self-report)	1.1	0.9	0.1	0.432	202	204	0.6	0.7	-0.0	0.849	218	216
Justified absences from school in the past month	0.2	0.1	0.1	0.181	202	204	0.1	0.2	-0.1	0.400	218	216

Source: Soft skills and self-reported absences: Student baseline survey data, April/May 2019 (TTH) and November/December 2019 (FM and MS). Test scores: MASSAR data for the 2017–18 school year. Justified absences: school administrative records data, May 2018 (TTH) and November/December 2019 (FM and MS).

Notes: For each BFI (Big Five Inventory) soft skill, scores reflect the average level of agreement (on a scale of 1 to 5, with 1 being strongly disagree and 5 being strongly agree) across 8–10 statements related to that soft skill. The minimum score is 1 and the maximum score is 5. For grit, the score reflects the average level of agreement (on a scale of 1 to 4, with 1 being almost always to 4 being almost never) across 5 statements. The minimum score is 1 and the maximum score is 4. A higher score means that the student exhibits that soft skill or grit to a greater extent. Test scores are out of 20. Columns A, B, F, and G present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

Table B.21 presents findings for student progression, repetition, and dropout rates by gender. Among lower secondary male students, there are two statistically significant differences between the two groups—7th grade progression rates (significant at the five percent level) and 7th grade repetition rates (significant at the 10 percent level). There is only one statistically significant difference (at the 10 percent level) between female students in both groups, which is about what we would expect due to chance. At the upper secondary level, there are three statistically significant differences between male students in the treatment and control groups, all of which are significant at the five percent level. These differences are for progression, repetition, and dropout rates for 10th grade students. Among upper secondary female students, there are five statistically significant differences. Differences between 10th grade progression and dropout rates and 11th grade dropout rates for female students in both groups are significant at the five percent level. Differences for 11th grade progression and repetition rates between the two groups of female students are significant at the one percent level.

Table B.21. Student progression, repetition, and dropout rates, by treatment status and gender

	Male						Female					
	Mean				Sample size		Mean				Sample size	
	Treatment (A)	Control (B)	Difference (A-B)	p-value (C)	Treatment (D)	Control (E)	Treatment (F)	Control (G)	Difference (F-G)	p-value (H)	Treatment (I)	Control (J)
Lower secondary students												
Progression (percentage)												
7th grade students progressed to 8th grade for the 2018–19 school year	55.2	58.5	-3.3**	0.036	10,124	42,591	76.7	78.5	-1.8	0.170	7,730	33,687
8th grade students progressed to 9th grade for the 2018–19 school year	63.6	64.9	-1.3	0.437	7,333	32,694	78.7	80.3	-1.6	0.186	6,525	30,225
9th grade students progressed to 10th grade for the 2018–19 year	48.9	50.2	-1.3	0.524	7,185	34,449	56.8	58.8	-1.9	0.278	7,477	34,410
Repetition (percentage)												
7th grade students repeating the grade for the 2018–19 school year	28.8	26.6	2.1*	0.052	10,124	42,591	14.5	12.9	1.6*	0.099	7,730	33,687
8th grade students repeating the grade for the 2018–19 school year	22.6	22.3	0.3	0.827	7,333	32,694	12.4	11.7	0.7	0.465	6,525	30,225
9th grade students graduating for the 2018–19 school year	32.2	32.4	-0.2	0.899	7,185	34,449	27.2	26.7	0.4	0.690	7,477	34,410
Dropout (percentage)												
7th grade students dropping out during or between school years	19.0	18.0	1.0	0.252	10,124	42,591	10.3	10.1	0.2	0.781	7,730	33,687
8th grade students dropping out during or between school years	16.1	15.2	0.9	0.298	7,333	32,694	10.4	9.5	0.9	0.113	6,525	30,225
9th grade students dropping out during or between school years	20.7	19.3	1.3	0.249	7,185	34,449	17.3	15.7	1.5	0.183	7,477	34,410

	Male						Female					
	Mean				Sample size		Mean				Sample size	
	Treatment (A)	Control (B)	Difference (A-B)	p-value (C)	Treatment (D)	Control (E)	Treatment (F)	Control (G)	Difference (F-G)	p-value (H)	Treatment (I)	Control (J)
Upper secondary students												
Progression (percentage)												
10th grade students progressed to 11th grade for the 2018–19 school year	82.0	77.8	4.2**	0.011	4,392	21,004	91.1	89.3	1.8**	0.046	4,522	22,914
11th grade students progressed to 12th grade for the 2018–19 school year	80.2	77.8	2.5	0.130	4,296	20,245	90.9	88.2	2.7***	0.000	4,381	21,669
Repetition (percentage)												
10th grade students repeating the grade for the 2018–19 school year	10.5	13.2	-2.7**	0.017	4,392	21,004	4.6	5.4	-0.8	0.246	4,522	22,914
11th grade students repeating the grade for the 2018–19 school year	12.2	13.8	-1.6	0.207	4,296	20,245	4.7	6.3	-1.5***	0.007	4,381	21,669
Dropout (percentage)												
10th grade students dropping out during or between school years	9.0	11.1	-2.1**	0.025	4,392	21,004	5.1	6.2	-1.1**	0.018	4,522	22,914
11th grade students dropping out during or between school years (percentage)	8.8	10.0	-1.2	0.108	4,296	20,245	5.1	6.5	-1.4**	0.010	4,381	21,669

Source: MASSAR data for the 2017–18 and 2018–19 school years.

Notes: Progression and repetition rates capture whether or not the student progressed to the next grade level or repeated the same grade level for the 2018–19 school year. Dropout rates include students who dropped out during or between the 2017–18 and 2018–19 school years. Columns A, B, F, and G present regression-adjusted means that account for the strata used in random assignment and for clustering at the school level.

***/**/* Difference between treatment and control group means is statistically significant at the .01/.05/.10 level.

This page has been left blank for double-sided copying.

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

Mathematica®
Progress Together

mathematica.org website

EDI Global, a Mathematica Company

Operating in Tanzania, Uganda, Kenya, Mozambique, and the United Kingdom