FINAL REPORT

Evaluation Design Report for the Secondary Education Activity of the Morocco Education & Training Project

July 31, 2018

Paolo Abarcar
Emilie Bagby
Galina Lapadatova
Caroline Lauver
Audrey Moore
Matt Sloan

Submitted to:
Millennium Challenge Corporation
1099 14th Street NW
Washington, DC 20005
Project Officer: Ryan Moore
Contract Number: MCC-13-BPA-0040 (CL-002)

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Matt Sloan
Reference Number: 50341.01.210.032.000

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

ACRONYMS

AREF	Regional Academy of Education and Training (Académies Régionales de l'Education et la Formation professionnelle)
EMIS	Education Management Information System
ERR	economic rate of return
GoM	Government of Morocco
ICC	intra-cluster correlation
ICT	information and communication technology
IRB	institutional review board
ISIM	Integrated School Improvement Model
ITT	intent-to-treat
LS	lower secondary
MASSAR	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
MENFP	Ministry of National Education and Vocational Training
MIAES	Modèle Intégré d'Amélioration des Etablissements de l'Enseignement Secondaire
O&M	Operations and Maintenance
PEI	Integrated School Project
PISA	Program for International Student Assessment
RCT	randomized control trial
SIP	school improvement plan
STEP	Skills Towards Employability and Productivity Program
TIMSS	Trends in International Mathematics and Science Study
ToR	terms of reference
ToT	treatment on the treated
US	upper secondary
WASH	water, sanitation, and hygiene

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

CONTENTS

TABLES

FIGURES

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

I. INTRODUCTION

Youth unemployment is currently a global policy issue, with North Africa having one of the highest youth unemployment rates in the world. In Morocco, the misalignment of youth skills with the needs of the labor market, coupled with a lag in job creation have been the drivers behind the high youth unemployment rate. More than half of the country's students leave the education system early or without proficiency in the skills required for the labor market, despite Morocco's high levels of access to education. Although Morocco's primary school gross enrollment ratio is 116 percent and its primary completion rate is near 99 percent (The Education Policy and Data Center 2015), the country currently lags in school completion rates at the secondary school level. Only 66 percent of lower secondary (LS) entrants and 34 percent of upper secondary (US) entrants complete the respective cycles. In addition, nearly three-quarters of secondary school dropouts are girls (The World Bank Group 2012). Morocco also underperforms on international student assessments and performance measures, ranking 73rd out of 76 countries in global education performance rankings, based on its performance in the 2011 Trends in International Mathematics and Science Study (TIMSS) where it performed in the bottom third in 8th-grade math and science (OECD 2015). These poor learning outcomes and misalignment of transferable skills contribute to a weak integration of graduates into the labor market. The unemployment rate for youths between the ages of 15 and 24 is disproportionately high—approximately 21 percent were unemployed in 2017 (The World Bank Group 2018). Furthermore, the inability to produce and integrate job seekers to meet the needs of a growing and changing economy is a potential constraint to the country's economic growth (MCC 2015; The World Bank Group 2012).

The Millennium Challenge Corporation (MCC) and the Government of Morocco (GoM) are addressing these education and employment shortfalls through the Morocco Employability and Land Compact, which was signed in November 2015 and entered into force in June 2017. The Compact consists of two projects – The Education and Training for Employability and the Land Productivity Projects. The Education and Training for Employability project consists of two Activities: The Morocco Secondary Education Activity and a Workforce Development Activity. Mathematica is responsible for evaluating the Secondary education Activity that seeks to improve the quality and workplace relevance of secondary education through several interventions, which begin in 2017 and continue through 2022. These interventions include (1) school-based interventions, such as developing a more autonomous and performance-based school management, revising the secondary education pedagogy to help develop transferable skills in students, and improving the quality of infrastructure in secondary schools; (2) strengthening the national education policy environment and information systems to use assessment results to drive educational decision making; and (3) developing a system to operate and maintain school infrastructure nationwide. The goal of the reforms is to ensure that students acquire skills that improve their chances of transitioning into the Moroccan labor market.

MCC has contracted with Mathematica Policy Research to conduct an independent evaluation of the Morocco Secondary Education Activity. The evaluation design will use a mixed-methods approach to assess the effects of the activity and will include (1) a randomized controlled trial (RCT) of the school-based subactivity (including a cost analysis) and (2) an implementation study of all three subactivities.

We describe the proposed methodology for assessing the impact and contributions of the Morocco Secondary Education Activity in this evaluation design report. Chapter II presents a summary of the activity interventions as well as an overview of the program logic and a review of existing international literature on the impacts of similar interventions. Chapter III presents a detailed explanation of the evaluation design, providing a discussion of the evaluation questions, methods, and data sources for the study's primary outcomes. Chapter IV presents our evaluation administration and management plan, including obtaining institutional review board clearance, how data will be protected and reported, and the roles and responsibilities of the evaluation team members.

II. THE SECONDARY EDUCATION ACTIVITY

In this chapter, we provide an in-depth description of the Secondary Education Activity and its three main subactivities. We discuss the program logic behind the interventions and include a review of existing literature on similar interventions and the ways in which our evaluation of these interventions can contribute valuable information to policy discussions.

A. Project description

Mathematica will evaluate the Secondary Education Activity, which is targeted at both lower secondary (LS) schools (grades 7 to 9) and upper secondary (US) schools (grades 10 to 12). The activity consists of three subactivities that seek to improve student acquisition of skills that are relevant to the private sector, which intends to create a more employable workforce:

1. **Integrated School Improvement Model (known as MIAES for its French name, Modèle Intégré d'Amélioration des Etablissements de l'Enseignement Secondaire)**.[1] The education consultant[2] will work with the Ministry of National Education and Vocational Training (MENFP) and with regional officials from the Regional Academy of Education and Training (AREF) to implement MIAES in 90 to 100 LS and US schools across the regions of Tanger-Tétouan-Al Hoceima, Fès-Meknès, and Marrakech-Safi. These school-level interventions include (1) developing and implementing school improvement plans (SIPs) to increase education quality and decrease gender and socioeconomic inequities at the school level through engagement of the community in decision making; (2) conducting capacity-building programs to improve school leadership and teachers' pedagogical delivery methods; and (3) improving school infrastructure (for example, classroom construction; water, sanitation, and hygiene [WASH] facilities). The GoM is also establishing a Partnership Fund to provide additional interventions and support to a subset of schools in the MIAES regions. [3] [4]
2. **Student Assessment and Education Management Information System (EMIS)**. This subactivity is comprised of two components that will be implemented nationwide. First, it provides support for the development and implementation of rigorous national and international student assessments as well as the use of the data to inform GoM's decision

[1] Some MCC documentation uses the ISIM acronym for its English name, while others use MIAES for its French name. We use MIAES because this is how it is known in Morocco.

[2] C2D has been contracted to carry out the work.

[3] GoM and MCC are still developing the design of the Partnership Fund so we do not have the details of its implementation at this time. We will not evaluate the Partnership Fund as a separate activity, but as part of the package of interventions received by the treatment schools. We will take into account that a subset of treatment schools is receiving this additional support when we interpret the results of the MIAES program. Mathematica will work with MCC, Millennium Challenge Account Morocco (MCA-M), and the Ministry of Education to mitigate any threats to the evaluation design when additional information on the Partnership Fund becomes available.

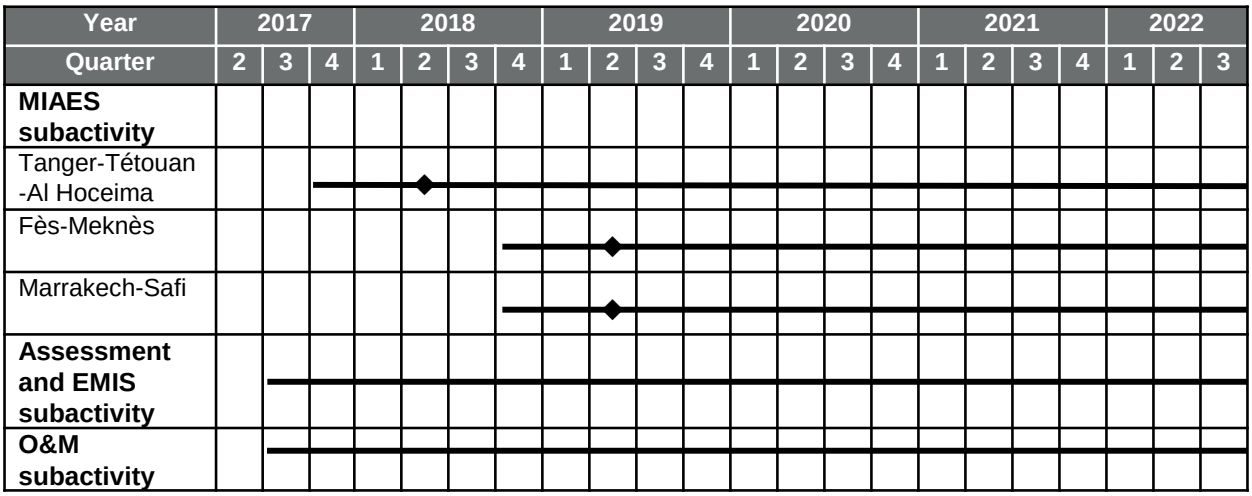
[4] In addition, activities related to addressing school-based violence may be included as a part of the Activity.

making.[5] Second, it includes technical assistance to improve the existing EMIS, called MASSAR.

3. **School Infrastructure and Equipment Operations and Maintenance (O&M)**. Interventions for this subactivity include (1) technical assistance to develop a new national approach to school infrastructure development, operations, and maintenance; (2) capacity building for regional actors engaged in the subactivity to support implementation of O&M; and (3) pilot testing of the use of performance contracts to maintain and operate school infrastructure and information technology.

The three subactivities will be implemented from 2017 through 2022. The consultant for MIAES began to pilot the participatory approaches of the Integrated School Project (PEI) (i.e. approaches to integrating school improvement plans) in six pilot schools in Tanger-Tétouan- Al Hoceima during the 2016–2017 school year, with the goal of identifying the best techniques for implementing the SIPs into the remaining target schools in Morocco. We understand that full implementation of teacher pedagogical training program and the school improvement plans, including capacity building of AREF, will begin in late 2017 for Tanger-Tétouan-Al Hoceima in selected schools. The interventions will roll-out in Fès-Meknès and Marrakech-Safi in 2018. The MIAES school infrastructure improvement component will likely roll out in the summer of 2018 in Tanger-Tétouan-Al Hoceima and in the summer of 2019 in Fès-Meknès and Marrakech-Safi. Planning for the EMIS and the O&M subactivities began in 2017; however, it is not clear when the interventions will reach the schools. We will closely monitor implementation and will revise the evaluation plan as necessary. Compact closeout will occur in the third quarter of 2022. Figure II.1 presents the implementation timeline.

Figure II.1. Implementation timeline



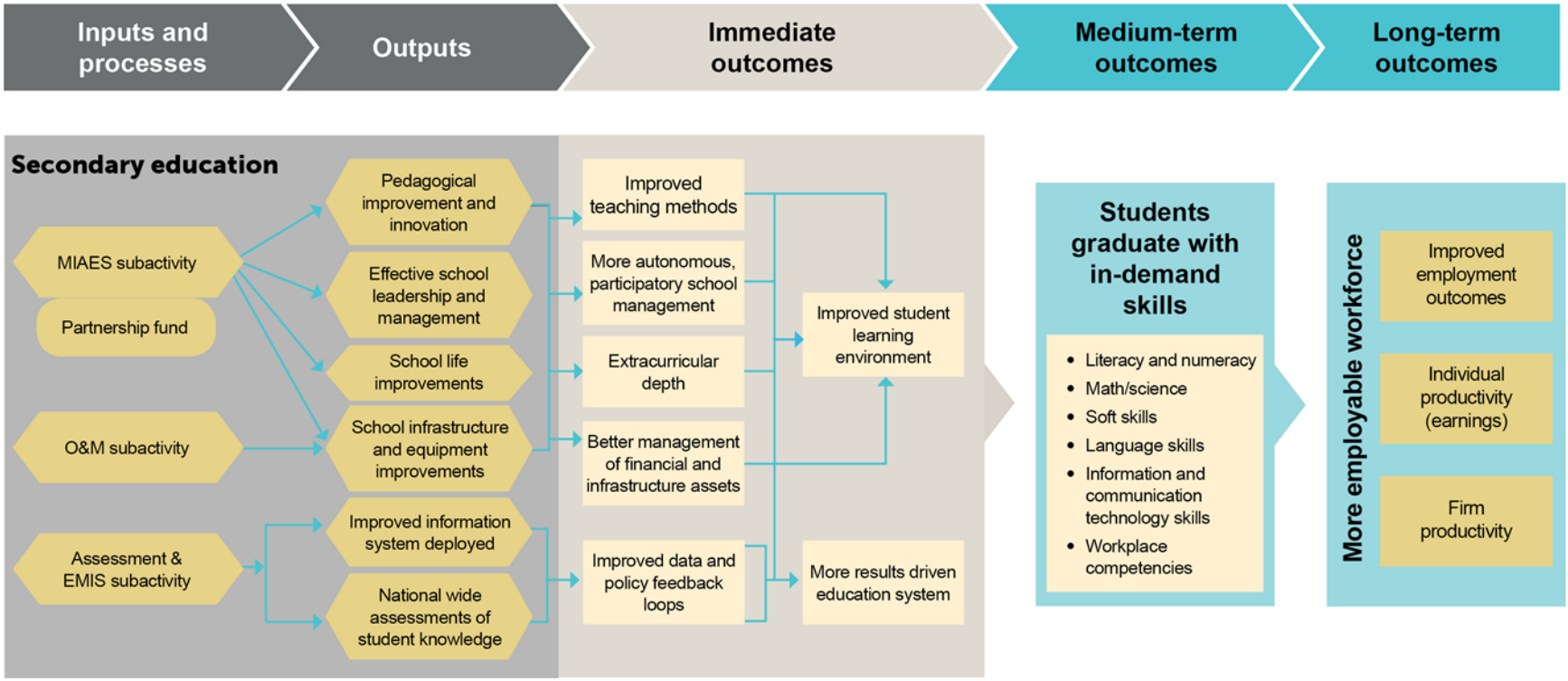
Note: A pilot program of the MIAES began in six schools in Tanger-Tétouan-Al Hoceima during the 2016–2017 school year and is not noted here. The black line indicates the timeline of implementation. The ◆ indicates the expected timing of the start of implementation of the school infrastructure component for the MIAES.

[5] We understand that GoM plans to participate in the PISA in 2018.

B. Program logic

The program logic presents a series of (hypothesized) causal links among program inputs and immediate, medium-term, and long-term outcomes that support the activity's overarching goal of enhancing access, appeal, quality and relevance of education in Morocco (Figure II.2). Each of the links in the program logic represents an assumption by the program designers about how activities will affect the compact's beneficiaries and stakeholders—which include students, teachers, school administrators, businesses, and policymakers in relevant ministries and centers.

Figure II.2. Program logic for the Secondary Education Activity



Source: MCC (2016), as modified by Mathematica Policy Research.

The MIAES subactivity is founded on an integrated school model that addresses multiple aspects of the school environment to improve student engagement and help them learn workplace-relevant skills. The MIAES supports and enhances school leadership, invests in improving the school environment, engages the community in decision making about the school through the development of school improvement plans, and prepares teachers to use active and student centered teaching methods. If successful, this package of interventions is expected to improve teaching methods, contribute to a more autonomous and participatory school management, increase extracurricular depth of the school program by providing additional activities or classes for students, and lead to a better learning environment. These improvements will lead to higher quality education, increased enrollment and retention of students, and improved learning outcomes. Teachers will be expected to improve their use and delivery of soft skills, which will help students develop workforce-relevant skills that will facilitate their transition into the workforce. Our literature review (Chapter II, Section C) show that these are reasonable assumptions from the school and educator component of the hypothesized causal chain.

The assessment and EMIS subactivity seeks to improve the ways in which GoM rigorously assesses the quality of education and improves the decision making process at all levels of the educational system. First, feedback for teachers on student performance through participation in nationwide assessments of student knowledge helps them understand the skills their students need to strengthen in the classroom. Second, results from assessments help policymakers understand overall student performance and develop capacity-building programs for teachers to improve weak content areas. Finally, accurate and up-to-date EMIS data assist policymakers in allocating resources to support the education system from top to bottom, which ensures that schools can retain students and help them learn. These contribute to a more results driven education system. Although the literature establishes that developing and using a results-driven system improves student learning, the key to success includes a strong focus on training teachers to use the assessment results to adapt their pedagogical delivery and support students who show poor results. It is also important for school directors and policymakers to clearly communicate results to communities so that people in the communities can hold schools accountable for results.

The O&M subactivity aims to improve the maintenance of school infrastructure nationwide, and will complement the infrastructure improvements under the MIAES. The improved O&M practices expected as a result of the O&M interventions should lead to better management of financial and infrastructure assets as well as to a better learning environment, and thus contribute to improved student retention and learning.

Overall, the program's theory of change suggests that if MCC invests in the combination of activities under the Secondary Education Activity, then students will acquire skills demanded by employers, such as literacy and numeracy, math and science, soft skills (e.g. perseverance, problem-solving, time management), language, information and communication technology (ICT), and other workplace-relevant skills. The assumption is that lack of these skills is the binding constraint to employability. When students receive instruction that facilitates development of such in-demand skills in schools that encourage retention of students, they are more likely to graduate from secondary school and are better prepared to transition into the

workforce at graduation. The improvement in the relevance of skills will contribute to workforce productivity and earnings.[6]

C. Literature review

Many donors and governments have launched secondary education reforms to improve teachers' abilities to integrate workforce-related skills into the core secondary school curriculum, improve school leadership and management, and create school improvement plans that engage communities in making decisions at the school level. The goal of these types of reforms is to help secondary school students remain in school longer, attain appropriate transferable workforce-related skills, and obtain better employment in the long term. To provide context for our evaluation of the Secondary Education Activity in Morocco, we reviewed the existing evidence on the effects of various types of education programs. In this section, we organize our review of the literature by the key outcomes and interventions of the activity. These include (1) student skills, employment outcomes, and earnings; (2) teacher training and investments in pedagogical innovations; (3) school-based management and student outcomes; (4) school infrastructure; and (5) data for decision making. For a situational analysis of the binding constraints to growth in Morocco and further discussion of the Moroccan context for the different types of planned activities, please refer to MCC's Constraints Analysis (MCC 2015) and MCC's Program Design Document (MCC 2016).

1. Student skills, employment outcomes, and earnings

It is well established in the literature that growth in individual productivity can be driven by improvements in school attainment and the development of cognitive skills (Groh et al. 2015; Hanushek and Woessmann 2007; Heckman et al. 2006; Heckman and Kautz 2012). Improving the quality of education can reduce the number of dropouts in schools, improve educational attainment and learning, and lead to better labor market outcomes (Hanushek and Woessmann 2007). Ibarraran et al. (2014) found that wages can be further improved when educational programs collaborate with the private sector because such collaboration often exposes students to the types of skills employers seek and gives the students experience (that is, internships, job shadowing) with using these skills.

Though it is a relatively nascent area of research in the economics literature, soft skills have also been posited to play a vital role in a range of life outcomes (Deming 2017). 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) suggested that soft skills could rival IQ in predicting educational attainment, success in the labor market, health, and even criminality. Heckman et al. (2006) found that latent noncognitive skills increased wages by directly affecting productivity and by indirectly affecting an individual's schooling and work experience. However, according to Blattman and Ralston (2015) and Rankin et al. (2015), there is currently a lack of rigorous evidence on the effectiveness of soft skills training, especially in less developed countries. There is little research that shows the impact that educational interventions have on the acquisition of soft skills (Guison-Dowdy [2012], as cited in Burnett

[6] It is however possible that acquiring workplace-relevant skills does not translate to better employability or earnings. If the binding constraints to employment are not the lack of worker skills, but something else, then the Activity may not succeed in achieving its ultimate goal of higher employment rates for youth.

and Jayaram [2012]) and their impact on 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) found that a female youth employment program in Jordan that was focused on building soft skills showed no significant impact on employment. However, the authors noted that the short duration of the teacher training program limited teacher uptake of knowledge and affected the ability of the teachers to support girls in learning the skills.

2. Teacher training and investments in pedagogical innovations

The literature suggests that support for teachers is critical to the continual development and effective use of pedagogical delivery skills. The literature also presents rigorous evidence showing a positive relationship between the professional development of teachers and students' performance. Evans and Popova (2015) looked at six of the most recent reviews of the impact of teacher professional development on student learning (Conn 2014; Glewwe et al. 2014; Kremer et al. 2013; Krishnaratne et al. 2013; McEwan 2015; Murnane and Ganimian 2014). These reviews examined the interventions that improved learning outcomes in low- and middle-income countries at both the primary and secondary levels. The results of this analysis showed that high quality teacher training interventions appear to be consistently effective at improving learning outcomes. However, the authors also noted that the effectiveness of a teacher training program depended upon the way it was implemented in the field. Muralidharan and Sundararaman (2010) noted that providing only general instruction or generic training to teachers tended to be ineffective, while ensuring that training programs tailored training to the skill levels of teachers showed strong effects on learning outcomes (Murnane and Ganimian 2014). The authors recommended providing more frequent visits of pedagogical advisors, ensuring that the support staff focused on skill development rather than classroom management, using teacher circles, and assisting with the development of scripted lessons. Other effective aspects of teacher training interventions highlighted in Evans and Popova (2015) included the provision of instructional materials as well as training teachers in the use of these materials (Krishnaratne et al. 2013; McEwan 2015).

To improve student learning outcomes through teacher-level interventions requires sustained and intensive professional development and support (Darling-Hammond et al. 2009). Structured pedagogy programs, which typically address several constraints to learning, have shown positive effects on student learning outcomes. However, these programs have to provide the right amounts (that is, implementation dosage) of the interventions to achieve results. Examples of implementation dosage in education include the amount of training that coaches and teachers receive in preparation to deliver a new pedagogical method, the amount of time coaches or pedagogy advisors spend working with teachers on the delivery of a scripted lesson, or the amount of time teachers spend receiving training on the use of new materials (Wasik et al. 2013). The existing research on dosage clearly states that one dose of an intervention is usually not enough (Boller et al. 2004; Joyce and Showers 1980; Winton and McCollum 2008). For example, holding a one-day workshop for teachers does not provide the necessary depth of understanding and is usually insufficient (that is, too infrequent, thus not reinforcing learning) to affect teacher learning or to change and improve long-term classroom practices (Boller et al. 2004; Raikes et al. 2006; Winton and McCollum 2008). However, there is little evidence that tells us the right dosage of teacher training that is required to move student learning in a positive

direction. The context, type of intervention, and dosage all matter to achieve significant effects on learning outcomes, but the programs that demonstrated the most success all had strong elements of teacher coaching and direct pedagogical support (Moore et al. 2017). Conn (2014) supported the importance of ongoing coaching during and after training workshops by showing that pedagogical interventions involving long-term teacher mentoring or in-school teacher coaching produced a sizeable (although not always significant) effect on student learning.

3. School-based management and student outcomes

There is limited but positive literature on the links between school-based management and improvements in student learning. Kirst et al. (2005) found that schools in California that took responsibility for student achievement used assessment data to improve teacher instruction, provided coaching and support to struggling teachers, and ensured that the availability of instructional resources had higher academic performance index scores than schools where the director did not prioritize these practices. Yamauchi (2014) found that a school-based management program in the Philippines that provided funding for school improvement plans increased the average national achievement test scores substantially over the course of three years. The school-based management grants were awarded based on the relevance of the school improvement plans and annual improvement plans that schools developed and submitted, giving schools an incentive to proactively and independently manage themselves. Carr-Hill et al. (2014) found that decentralization can have significant impacts on student dropout and retention rates across low- and middle-income countries because parents and the community have more input in ensuring that children stay in school. However, school-based management interventions may also take a long time to affect change. A World Bank review of the literature by Bruns et al. (2011) found that although school-based management and accountability interventions in Latin America achieved positive results on attendance, repetition and failure rates, dropout rates, and test scores, it took nearly five years to observe meaningful changes in test scores, retention, and completion. Similar results were found in the Philippines, where school-based management interventions showed positive effects on test scores. However, it took more than three years to begin to see changes from that program (Yamauchi 2014).

4. School infrastructure

There is some evidence 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) all have suggested that school infrastructure expansions and improvements can have a positive impact on student enrollment and achievement.[7] 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 types of infrastructure improvements induce students who have either dropped out of the system or who never enrolled to enroll and potentially stay longer in school. 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).

[7] See also Bagby et al. (2017), Campuzano et al. (2016), and Bruce et al. (2017).

5. Data for decision making

The evidence on how effectively teachers, school directors, and Ministry of Education staff use student outcome data in their decision making is mixed. Recent studies of initiatives in India, Kenya, Tanzania, and Uganda, where nonprofit organizations implemented citizen-led assessments of children's basic reading and math skills, found that these assessments did not translate to improved student learning even though the tests increased awareness of problems in the education system. Lack of available resources and limited capacity of school officials to implement reforms ultimately hindered systematic action (Results for Development Institute 2015). The literature has also shown that student assessments and information systems alone do not often improve policy feedback loops. Improvement is more likely if the system also includes systemic change and incentives for the information to be used for improvement (Results for Development Institute 2015; Szekely 2011). The literature points out that each country usually employs a range of assessments in various formats and settings, but in many respects the assessment systems do not function optimally. There is insufficient attention given to training teachers to use assessment results effectively in the classroom to adjust their lesson plans or help lagging students. There is also little emphasis on training local, regional, and national education staff to develop the communication channels, feedback mechanisms, and protocols for utilizing the data that characterize data-driven organizations (Braun and Kanjee 2006). In the end, the key to linking student assessments with improved student and system performance lies in how educators and policymakers understand and use the data. Without a focus on this aspect of training and data use, the link between assessments and student outcomes becomes tenuous.

D. Policy relevance of the evaluation

Our review of the literature shows that the evaluation of the Secondary Education Activity of the Morocco Education Training Project has the potential to contribute to gaps in the understanding of how and why specific education interventions related to school management, school infrastructure, teacher training in student-centered pedagogy, and improved education data at a national level can result in students' improved educational attainment, cognitive and soft skills learning, and employability. Although the literature establishes that interventions such as teacher training, pedagogical improvements, and strong school management can have impacts on student performance, the size of the impact depends upon the dosage, duration, and fidelity of implementation, as well as the local context. Which of these matters most is unclear. The literature also establishes that school-based management and school infrastructure improvements can contribute to improvements in student outcomes over time, though the magnitude and the timing of the effects is not clear. The literature also demonstrates that it is important for teachers to have consistent and ongoing support as well as strong feedback loops, ensuring that teachers can request support, receive assessment data, and be taught how to use that data to improve their students' learning. It is critical for policymakers to use the feedback loops to share data that inform how current and new policies are developed and implemented at all levels of the system. However, the literature does not reach a consensus on the most effective way to create these important information systems. In addition, the evidence on the effects on soft skill acquisition is sparse. There is little evidence on which pedagogical innovations help students develop soft skills and whether the school management approach has any effect on how soft skills are acquired in schools.

A rigorous evaluation of the MIAES package of activities, which will be complemented by a qualitative study of the MIAES, assessment and EMIS, and O&M subactivities, has the potential to make significant contributions to both policymaking in Morocco and to the literature more broadly. First, the studies will allow MCC and the GoM to rigorously attribute effects to the package of interventions in the MIAES 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 additional 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 the facilitators of and the barriers to that prevent the use of data and provide insights into how trained staff have used information effectively to help improve the education system at all levels.

III. EVALUATION DESIGN

In this chapter, we describe our proposed design for the evaluation of the Morocco Secondary Education Activity, which includes an RCT and a qualitative study. We list the research questions that the evaluation addresses and provide a brief overview of the proposed evaluation design. We then describe each of the two components of the evaluation in more detail, including the study samples involved, sample sizes, data sources, and analytical approach. Finally, we discuss data quality, the project timeline, and dissemination plans, as well as the limitations and challenges of the evaluation.

A. Research questions and overview of the evaluation

The research questions that we will study through a rigorous evaluation and a qualitative study are related to understanding (1) impacts of the different intervention activities on students, teachers, and schools and (2) how the different intervention activities cause change in the experiences and behaviors of these actors as well as in the education system in Morocco. To answer these questions we will conduct a mixed-methods evaluation that includes an RCT of the MIAES and a qualitative study of all three subactivities. The mixed-methods approach will allow us to bring together quantitative and qualitative methods to strengthen the validity and reliability of our findings.

Table III.1 lists the research questions for the study (numbered 1 to 14). The questions are subdivided by activity and by the unit of measurement (student, teacher, school, or system). The table also links the evaluation questions to the type of study (RCT or qualitative study).

The RCT will provide rigorous estimates of the causal impact of the MIAES subactivity on student outcomes such as enrollment and learning; teacher practices and attendance; and the quality of school infrastructure (research questions 1 to 5 and 8). We will quantify impacts by randomly assigning which schools receive the MIAES and by comparing outcomes in schools assigned to receive the MIAES versus schools assigned to continue as usual. We will conduct a longitudinal student survey, a student skills assessment, a teacher survey, direct classroom observation of teachers, a school director survey, and a school infrastructure checklist at baseline and in the follow-up period. Because MCC is interested in distinguishing between impacts that occur for lower and upper secondary schools independently, we will separately analyze the impacts for both types of schools in the RCT.

We will supplement the RCT with a qualitative study that describes how change occurred through the MIAES. The analysis will include in-depth qualitative data to help us understand changes in teaching practices; how school leadership and management contributed to improved learning at the school; how autonomy and accountability might have improved in participating schools; and the role that parents and the community play in implementing the school improvement plans (research questions 4, 6, 7, 9, and 10). This approach will deepen our understanding of the research questions related to the RCT. We will rely on focus groups with students, parents, and teachers and in-depth interviews with school directors in 2020, three years after the initial rollout of the intervention activities and a year before to the compact's closeout. The qualitative study will also seek to understand the implementation process and the potential effects of the EMIS and O&M subactivities (research questions 11 and 12), which will be

implemented in secondary schools nationwide. We will conduct key informant interviews with Ministry of Education officials, in addition to desk reviews of program documents. We will use qualitative analysis methods to understand how improved data and policy feedback loops might have led to a performance-driven education system, including better management of financial and infrastructure assets, and ultimately to an improved student learning environment. The analyses will allow us to understand the sustainability of the interventions within Morocco (research question 13).

In addition, using our estimates of the effect of the Secondary Education Activity from the RCT, we will also conduct a cost-benefit analysis to calculate the economic rate of return (ERR) for the interventions (research question 14).

Table III.1. Research questions and evaluation design

		RCT	Qualitative study
MIAES 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 MIAES 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? Please provide examples.		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
Assessment and EMIS subactivity			
System	11. How did the interventions contribute to improved student assessment, data, and policy feedback in the EMIS system, 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 and scale the interventions under the Secondary Education Activity?		X
System	14. To what extent are the interventions under the Secondary Education Activity cost-effective? (E.g. Can the Ministry of Education financially sustain the interventions? What is the economic rate of return to the beneficiaries?).	X	

B. RCT to evaluate the MIAES subactivity

In this section, we discuss our approach for the rigorous evaluation of the MIAES subactivity. We begin by explaining the process of conducting random assignment. We discuss the study sample, recommended sample sizes for data collection, and data sources. Finally, we end with an explanation of our analytical approach and our data analysis process.

1. Random assignment

The impact evaluation of the MIAES involves random assignment, which is the most rigorous way to estimate causal impacts. The method randomly assigns schools to a group that receives the intervention (the treatment group) and to a group that does not (the control group). Random assignment ensures that school, teacher, and student characteristics do not determine treatment status and that observable characteristics should be the same on average in both treatment and control groups prior to the intervention. 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 and the GoM committed to an RCT of the MIAES subactivity in 2016. Mathematica conducted random assignment of schools in the first region, Tanger-Tétouan-Al Hoceima, in December of the same year. Random assignment for the next two regions, Fès-Meknès and Marrakech-Safi, will proceed in early 2018. MCC, Millennium Challenge Account Morocco (MCA-M), and the MENFP selected these three regions to be nationally representative, in order to maximize the potential for learning from the evaluation and for scale-up of the program (MCC 2016).[8] MCC and the GoM decided that 57 LS schools and 27 US schools in these regions would receive MIAES and participate in the study. These numbers exclude 6 schools (3 LS and 3 US) that received the pilot program and therefore will not be included in random assignment.

Below we elaborate on the process we will use to identify eligible schools for random assignment within each region, to group eligible schools into strata, and to randomly select treatment schools among eligible schools within each stratum.

a. Selection of eligible schools for random assignment

The MENFP, MCA-M, and MCC will identify schools that are eligible to receive the MIAES and participate in random assignment. They will select up to four provinces in each of the three MIAES program regions. In these provinces, MCA-M and the MENFP screen all LS and US schools for eligibility. Schools are eligible as long as they (1) are not condemnable or slated for demolition, (2) do not have asbestos, or (3) are not undergoing rehabilitation (or have not yet reopened after a rehabilitation). Also, schools must not have structural problems and must

[8] In order to achieve national representation, 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.

have at least 288 students enrolled (50 percent of the built capacity of the smallest model school) to be determined eligible for the MIAES.[9]

For the first region, Tanger-Tétouan-Al Hoceima, the MENFP, MCA-M, and MCC selected the following four provinces for the MIAES: (1) Tanger, (2) Tétouan, (3) Larache, and (4) Chefchaouen. The same economic, educational, and geographic criteria that were used to balance national representativeness among chosen regions were also used to select the provinces. The MENFP and MCA-M determined the eligibility of schools to participate in random assignment through site inspections and verification of school records. Through this process, 180 schools in these four provinces were identified, of which 133 schools were deemed eligible to participate in the random assignment. Eighty-five schools were LS, 38 were US, and 10 had both levels in the same school.

In Fès-Meknès and Marrakech-Safi, the MENFP, MCA-M, and MCC have yet to finalize the list of provinces and eligible schools. They will do so in early 2018. Assuming that the composition of schools in these two regions is similar to Tanger-Tétouan-Al Hoceima, we expect that approximately 261 LS and 150 US will be eligible for random assignment across all three regions.[10]

b. Random assignment of schools

Once the selection of eligible schools is complete, Mathematica, in collaboration with MCC, MCA-M, and the MENFP, will proceed with random assignment. The process involves several steps including identifying strata, determining the number of schools to select by strata and conducting random assignment in a public lottery.

Grouping schools into strata. The first step is to group the lower and upper secondary schools by strata. We divide the schools according to their province, urban or rural status, and school type (whether LS or US). We then randomly assign eligible schools as follows:

- Eligible LS schools are randomly assigned to a treatment or control group, using strata defined by province and urban or rural status.
- Eligible US schools are randomly assigned to a treatment or control group, using strata defined by province (and by urban or rural status if feasible).

Table III.2 illustrates how this process occurred in Tanger-Tétouan-Al Hoceima. The table depicts the number of schools that we grouped in each stratum for this region. For example, it

[9] We have gathered this information from project design documentation from MCC.

[10] Including the 6 pilot schools, there were 88 LS schools and 51 US schools eligible to receive the MIAES in Tanger-Tétouan-Al Hoceima. Schools that have both LS and US levels in the same school are considered US schools for the purposes of random assignment. Thus, assuming that Fès-Meknès and Marrakech-Safi have the same number of eligible schools as Tanger-Tétouan-Al Hoceima, there will be 264 (= 88*3) LS schools and 153 (= 51*3) US schools. Subtracting the pilot schools leaves 261 LS schools and 150 US schools in total that are eligible for random assignment.

was necessary to combine urban and rural US schools into the same stratum within each province because there were only a few schools in these strata to conduct random assignment. For those schools that had both lower and upper secondary levels on the same premises and therefore would both be affected by the intervention activities, we consulted with stakeholders and agreed to consider these schools as US schools for the purpose of the random assignment and the analysis.[11] We expect to follow the same process in the other two regions.

Table III.2. Number of eligible LS and US schools in Tanger-Tétouan-Al Hoceima, by stratum

	LS schools		US schools
Province	Urban	Rural	Urban/Rural
Chefchaouen	3	9	11
Larache	13	7	8
Tanger	28	0	17
Tétouan	17	8	12
Total	**61**	**24**	**48**

Note: Urban and rural US schools within each province were combined into the same stratum. Ten schools that had both LS and US levels on the same premises were considered US schools for random assignment.

Calculating the number of schools to select for treatment. The second step is to calculate the number of schools to select for the treatment group in each stratum. Across all three regions, 90 to 100 schools will receive MIAES (for our calculations, we assume 90), divided roughly evenly across regions. In Tanger-Tétouan-Al Hoceima, 34 schools will receive the program, of which 6 are pilot schools that did not participate in random assignment. We follow the procedure described below.[12]

- First, the project stakeholders determine through consultation and discussion how many US and LS schools would be selected. Of the 28 schools to be selected to receive MIAES in Tanger-Tétouan-Al Hoceima through random assignment, 19 LS schools and 9 US schools would be selected, as determined by MCC, MCA-M, and the MENFP. This is roughly proportional to the number of each type of school that is eligible. We expect similar numbers in the other two regions.
- We then determine the breakdown of LS and US treatment schools by urban and rural area within each province. We divide the number of allocated treatment schools in each group—LS-urban, LS-rural, and US—across provinces in proportion to the number of eligible schools in each province. The breakdown is based on the proportion of the number of eligible schools in each secondary school level. For example, in Tanger-Tétouan-Al Hoceima, 72 percent of all eligible LS schools are in an urban area. Hence, 14 urban LS

[11] The evaluation will occur after three years of exposure to the intervention. Therefore, students in these schools would only have been exposed to either LS or US education. However, these schools will be treated as US schools in the analysis since the students we are tracking will be in US at follow-up data collection. We will conduct additional analyses that include LS students from these schools (while also weighting for the probability of being included).

[12] We present further details on this procedure in Appendix B.

(that is, 72 percent of 19) and 5 rural LS schools would be assigned to the treatment group (after rounding). Because there were limited numbers of rural US schools in Tanger-Tétouan-Al Hoceima, we did not separate those schools into a different stratum. We expect this to be the case in the other regions as well.

The number of schools to select for the treatment group for each stratum in Tanger-Tétouan-Al Hoceima is shown in Table III.3. We selected one LS-urban school in Chefchaouen, three LS-urban schools in Larache, six LS-urban schools in Tanger, four LS-urban schools in Tétouan, and so on. The process ensures that treatment assignment probabilities are similar across provinces within urban and rural areas and in LS or US schools. We will follow a similar process in the other regions once eligible provinces and schools are identified.

Table III.3. Number of LS and US schools in each stratum selected for the treatment group in Tanger-Tétouan-Al Hoceima

	LS schools		US schools
Province	Urban	Rural	Urban/Rural
Chefchaouen	1	2	2
Larache	3	1	2
Tanger	6	0	3
Tétouan	4	1	2
Total	**14**	**5**	**9**

Conducting lotteries for schools in a public ceremony. The final step is to hold a public lottery for random assignment. We will organize a public ceremony with school authorities and representatives of the MENFP to ensure that the random assignment process is transparent in all regions. We will conduct random assignment by drawing wooden blocks from bags. Each school is assigned a block. The blocks are then placed in bags that represent the strata. School authorities take turns selecting wooden blocks. The schools chosen in the ceremony receive the MIAES interventions.

We suggested and conducted a public ceremony in Tanger-Tétouan-Al Hoceima because such an event can bring together a wide range of school officials, including the Minister of Education, parents, teachers, students, and other community members (see Figure III.1). Publicly involving stakeholders in the ceremony engages school authorities in verifying whether the correct schools are marked to be selected for the treatment group. A public ceremony also generates understanding among community members that random assignment is a fair process of allocating scarce resources.

Figure III.1. The public selection ceremony in Tanger-Tétouan-Al Hoceima

Source: Ryan Moore, Millennium Challenge Corporation

2. Study sample for the RCT

All schools eligible to receive MIAES comprise the study population, and will comprise the study sample for the rigorous evaluation of MIAES. That is, all schools involved in random assignment will serve as our study sample for the RCT in analyses using administrative data. Similarly, all students and teachers in the study schools will comprise study sample in analyses using administrative data. A subsample of these schools, teachers and students will comprise the study sample for analyses using survey data. The following subsections discuss the sample for each of these groups.

a. Schools

All schools eligible for random assignment are a part of the study and will be included in the analyses using administrative data, and a randomly selected subsample of schools will be visited to collect survey data. We will use administrative data from the EMIS for the universe of schools that are eligible for random assignment to conduct impact analyses at the student, teacher, and student levels. In addition, we will draw a sample of each to survey for the impact analyses. We will survey all treatment schools, and stratify by using the same strata used for random assignment to draw a sample of control schools proportional to the number of schools in each stratum. We will also sample students and teachers for interviews from each school that is sampled.

b. Student cohorts

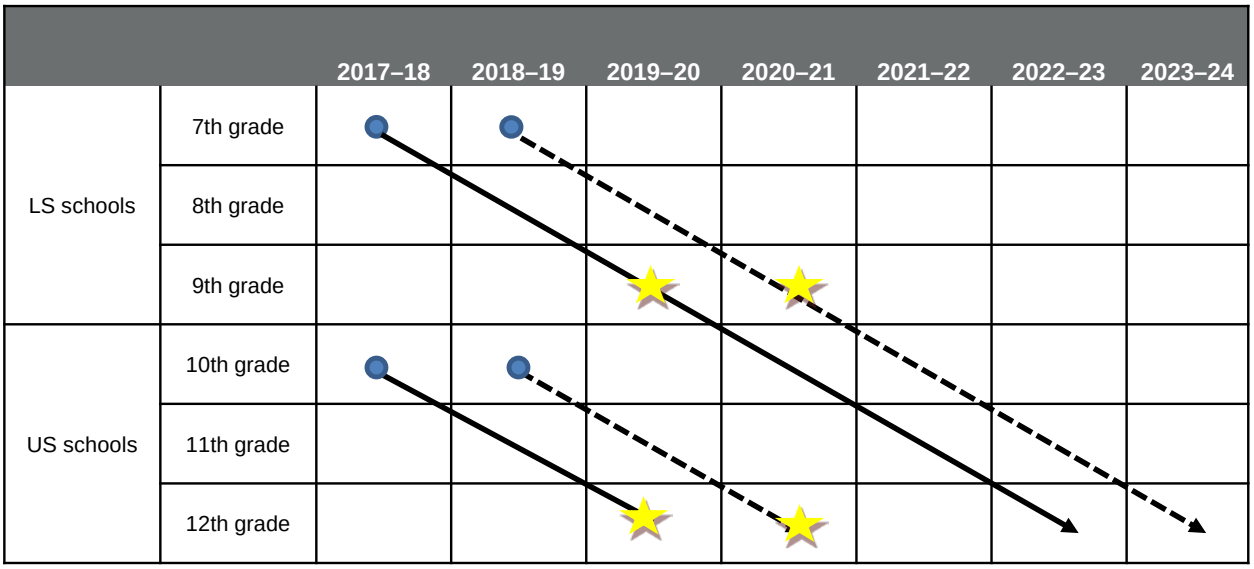
Our student sample for the RCT will be the entering cohort of LS and US students in the school years right after the schools' random assignment occurs. These cohorts include the 7th- and 10th-grade students in Tanger-Tétouan-Al Hoceima in the 2017–2018 school year and the 7th- and 10th-grade students in Fès-Meknès and Marrakech-Safi in 2018–2019. These students represent those who will be exposed to program activities as they go through each year of their respective levels of secondary school and will have up to three years of exposure to the program at endline. However, if there is a delay in the rollout of the interventions to schools, then the selection of these cohorts may change.

Focusing on students who begin schooling right after random assignment minimizes the possibility of student sorting over time, which could compromise the random assignment design. If students have information about the intervention that affects their choice of school, then the treatment and control groups would not be equivalent. It is feasible for students to choose which school they attend through their choice of study focus. If the closest secondary school (typically US level) does not offer the career path of interest, then a student may choose to attend a different school. One way students sort is when motivated and higher-ability students choose to enroll in treatment rather than control schools because they learn that treatment schools receive additional resources. This type of sorting may undermine our impact estimates because it works against random assignment by reducing the comparability of students in treatment and control schools. Focusing on the entering cohort right after random assignment minimizes the available time that students have to sort into different schools.[13]

We recommend following a longitudinal sample of LS and US entrants in treatment and control schools for data collection, because estimating learning impacts only from students who remain in school at the time of follow-up (such as in a cross-sectional sample) may produce biased estimates. Bias would arise primarily because outcomes would not be measured by the school system for students who drop out, and those students may be different from students that remain in school. For example, if treatment schools can retain more lower-ability students than control schools, then we might find that the interventions had a negative impact on test scores if we only focused on enrolled students at endline. The lower test scores could be attributed to the fact that a high number of lower-ability students in treatment schools (who would have otherwise dropped out) lowered the school's performance on assessments. Longitudinal tracking of students allows us to determine that changes in student outcomes result from actual learning, rather than from compositional changes in schools that may arise from the influx or exit of students. Figure III.2 illustrates our longitudinal approach.

[13] As an alternative to the study sample described above, we considered choosing the next cohort of entering students, the 7th- and 10th-grade students in Tanger-Tétouan-Al Hoceima in 2018–2019 and the same cohort in Fès-Meknès and Marrakech-Safi in 2019–2020. These students would receive all program activities including the school infrastructure interventions, which begin later than the rest of the MIAES activities. The impact estimates from this cohort would, therefore, capture the impact of all activities for a longer period of exposure. However, using this alternative sample increases the risk of student sorting—given the more extended period after random assignment that students are given to choose schools—which could lead to biased impact estimates. The visibility of school construction in treatment schools may also exacerbate student sorting if households see improvement in these schools and decide to send their children to these schools.

Figure III.2. The study sample will be the entering cohort of LS and US schools the school year immediately following random assignment



Note: The arrows represent the longitudinal sample of students whom we intend to follow for the evaluation. These students are the 7th- and 10th-grade students in Tanger-Tétouan-Al Hoceima in the 2017–2018 school year (represented by the solid black lines) and the 7th- and 10th-grade students in Fès-Meknès and Marrakech-Safi in 2018–2019 (respresented by the broken black lines). The dots show when we will conduct the baseline survey and the stars indicate when we will conduct the endline survey of these students.

We recommend the collection of baseline information from students in their schools and the collection of follow-up information from the same students after three years, just before they are due to graduate. If students drop out or move in the interim, we will attempt to find them to collect data at follow-up.[14] Focusing on these specific cohorts provides three distinct advantages:

1. It allows us to estimate impacts for students who will have exposure to program activities for the full cycle of secondary school. Unless students repeat a grade, three years at lower or upper secondary school represents the maximum period that students can receive the MIAES interventions at the given secondary level. After these years, students are no longer exposed to the interventions because they either move on to the next cycle or drop out of school (unless they decide to enroll in an upper secondary school that receives MIAES). Therefore, we argue that measuring learning at the end of these three years will provide an appropriate opportunity to estimate intermediate impacts on learning. Measuring learning impacts later is also possible, but it is unlikely that learning impacts will show up at a later point in time if effects do not manifest already after maximum exposure to the program. Measuring impacts before three years is also possible but only gives the partial effects of exposure to MIAES.

[14] The study sample will be the entering cohort of LS and US schools at the time of random assignment. We will explore using administrative data from the EMIS to identify students who leave the education system during the course of the evaluation, to determine how large of an issue it will be at endline.

2. It allows us to estimate effects independently for lower and upper secondary school.
3. It allows us to minimize attrition at follow-up data collection by interviewing students before they graduate from their respective cycle.

c. Teacher sample

Our teacher sample is will include teachers in treatment and control schools. At baseline, we will sample teachers of core subjects who are the target of the pedagogical training program. We will work with the Ministry of Education and C2D to identify the sample of teachers who are most likely to receive the interventions. In the treatment schools, this group represents the direct beneficiaries of teacher training. In the control schools, the teachers represent those who would have received training if their school had been selected to receive the MIAES. We anticipate that teacher turnover is relatively low in Morocco and therefore think that it will be relatively straightforward to collect data on the same sample of teachers over time. However, prior to finalizing our sampling approach for teacher selection, we will review MENFP EMIS data on teacher movement to determine if a longitudinal or cross-sectional sample of teachers is the most appropriate for this study given the potential issues of attrition and turn-over, particularly in rural schools.

3. Power calculations for sample sizes in the RCT

We calculated minimum detectable effects (MDEs) to determine the appropriate sample size of schools, teachers, and students in the sample. MDEs are the smallest impacts that the evaluation will be able to detect, given a particular sample size and conventional levels of statistical significance and power and are expressed in standard deviation units. The larger the sample size, the smaller the effects that can be detected. The goal is to determine sample sizes that balance the cost of data collection with the ability to detect impacts that are relevant for policymaking.

MDEs vary for different outcomes, and the size of impact that one might expect from a given intervention will vary depending upon the outcome. The literature indicates that a reasonable MDE to aim for in this context is 0.2 standard deviations for student outcomes. Such an effect size for student-level outcomes is within the range of impacts typically estimated in developing countries for education interventions that seek to improve learning outcomes. (Damon et al. [2015] is the most recent review study that compiles effect sizes found by various studies.) Pedagogical interventions, providing school inputs (such as textbooks or new infrastructure), and school governance interventions have rarely produced effect sizes larger than 0.3 standard deviations. We might, however, expect larger effect sizes than our benchmark of 0.2 standard deviations for teacher- and school-level outcomes. Teachers and school directors are direct beneficiaries to teacher training and school management interventions under the MIAES. Therefore, it is reasonable to expect that changes in their behavior are larger than what we would expect for student learning.

We discuss our recommendations for sample sizes for survey and administrative data below for our sample of schools, students, and teachers.

a. Sample sizes for survey data

We recommend collecting survey data from all treatment schools and from a sample of control schools to balance the cost of data collection with the statistical power gained for the study from each additional control school. Out of the projected 261 LS schools included in random assignment (57 of which will be selected for treatment), we recommend gathering survey data from a balanced sample of 57 treatment and 57 control schools. Out of the anticipated 150 US schools included in random assignment (27 of which will be selected for treatment), we recommend gathering survey data from a sample of 27 treatment and 27 control schools. We recommend collecting data on all treatment schools because collecting data from as many units of the smaller group of schools as possible (in this case treatment schools) maximizes statistical power. However, we recommend collecting survey data from the same number of control schools but not all control schools to limit the cost of data collection. The gains in statistical power from collecting data from all control schools would be low. We recommend randomly sampling the control schools for the survey from the population of control group schools proportionally by stratum (defined during random assignment), such that the survey sample contains the same number of treatment and control schools in each stratum.

We base our recommendation for the number of control schools to sample for the survey on the trade-offs between sample size, MDEs, and costs. In our calculations, we systematically varied the number of control schools and determined that surveying 57 LS and 27 US schools (that is, the same number as the treatment group) was optimal. In Appendix B, Figure B.1, we show how MDEs for student, teacher, and school outcomes vary with respect to the sample size of control schools for LS and US schools (holding the sample size for treatment schools fixed at 57 and 27, respectively, and holding the number of students and teachers surveyed per school fixed). Fewer schools lead to worse statistical power, but increasing the number of control schools above 57 LS and 27 US leads only to small improvements in the ability to detect effects. The costs of increasing the data collection sample size may not be justifiable.

We will sample students and teachers from the sampled schools to follow longitudinally in a survey. We recommend randomly sampling 15 students and six teachers per school.[15] When possible, we will ensure a gender balance in the selection of students and teachers.[16] To determine these numbers, we systematically varied the number of students and teachers in the survey sample to see how this affected MDEs (as shown in Appendix B, Figures B.2 and B.3). Choosing a sample size for students or teachers larger than this has little effect on MDEs for student- and teacher-level outcomes. One reason for this is because the sample size of the unit of random assignment (in this case, schools) influences the MDEs more than the unit of analysis (students). Surveying more students and teachers is unnecessary and costly, especially since we plan on tracking the same individuals over time. However, because of concerns about survey attrition at follow-up—especially for students—we recommend a minimum sample of 15 students and six teachers per school. We expect a significant proportion of students—34 percent in lower and 66 percent in upper secondary school (Education Policy and Data Center 2015)—to

[15] If there are less than 6 teachers in a school then we will interview all of them.

[16] In particular, we will determine if we will need to oversample males or females based on the cohorts in the study and their gender distribution.

leave school before endline. While we will attempt to follow all students in the sample, we do not expect to be able to track and interview all those who leave school. Therefore, we have assumed a survey attrition rate of 20 percent at endline for the student sample (which, depending upon the easy of following students over time, may be optimistic).[17] We discuss this issue in further detail in the challenges section found in Chapter II, Section H.

Our proposed sample sizes for survey data collection involve 1,710 students and 684 teachers in 114 LS schools (57 treatment and 57 control schools) and 810 students and 324 teachers in 54 US schools (27 treatment and 27 control schools), as shown in Table III.4. We show the MDEs for measuring impacts on students, teachers, and schools associated with these sample sizes for the LS school analyses and for the US school analyses. We believe that these sample sizes strike an appropriate balance between the cost of data collection and the precision needed to estimate project impacts. These sample sizes will allow the estimation of effects as small as 0.24 standard deviations for LS school impacts at the student level, which is within the range of impacts typically estimated from similar successful education evaluations.

Table III.4. MDEs for student-, teacher-, and school-level outcomes, using survey data with our recommended sample sizes

	Treatment sample size	Control sample size	MDEs for student-level outcomes (standard deviations)	MDEs for teacher-level outcomes (standard deviations)	MDEs for school-level outcomes (standard deviations)
LS school impact	57 schools 342 teachers 855 students	57 schools 342 teachers 855 students	0.24	0.27	0.44
US school impact	27 schools 162 teachers 405 students	27 schools 162 teachers 405 students	0.35	0.39	0.64

Note: MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. We based the total number of eligible schools on MCC program documents and data from the Tanger-Tétouan-Al Hoceima region, where data is currently available. In particular, we estimated the total number of schools by assuming that the number of eligible schools in the study areas is similar to the number of eligible schools found in the Tanger-Tétouan-Al Hoceima region. We expect to collect survey data for 15 students per school and 10 teachers per school. We assume attrition to be 20 percent for students and 10 percent for teachers at endline. We assume that the proportion of individual-level variance in the outcome explained by covariates for students and teachers is 0.40 and that the proportion of group-level variance explained by covariates is 0.30. We apply an intra-cluster correlation (ICC) of 0.25, as estimated from the TIMSS test score data for 8th-grade Moroccans in mathematics. We assume the same ICC for teacher outcomes.

We recognize that estimating impacts at the US school level (or for school-level outcomes in general) using survey data remains a challenge with this design. In Table III.4, we show that the MDEs for these impacts remain above 0.30 standard deviations for our recommended sample sizes. This means that even if there were quite large effects from the MIAES, we might not have the power to statistically detect them. Unfortunately, there is no clear solution to this problem, because the number of treatment schools cannot increase. Our analysis suggests that even

[17] We will attempt to monitor the number of school leavers using administrative data before endline to gauge the magnitude of the effort to follow them, and will work with the local data collection firm to identify ways of doing so.

collecting data on all control schools or on more students will not bring down these MDEs sufficiently to justify the cost of data collection. We discuss some strategies to factor this risk in our analysis in the challenges section found in Chapter III, Section H.

Even if the MDEs for impacts at the US or school-level are quite large, we believe that these effect sizes are still plausible. The MIAES may have large effects, given thethe package of interventions and the large amount of resources in each of the selected upper secondary schools. It is not unreasonable to expect effects of 0.3 standard deviations given the investments.

b. Sample sizes for administrative data

We expect to have access to administrative data from the full sample of secondary schools included in random assignment from Morocco's EMIS system. This includes information on our full study sample of students and teachers—that is, the entering cohort of LS and US students in the school year right after random assignment and the teachers of core academic subjects. We hope to be able to access information including student test scores from regional and national assessments, student enrollment and completion, teacher information such as attendance, and school enrollment and completion rates. If the ministry is willing to share these data, we would expect to use the data at little cost to the evaluation. We assume that we will be able to track outcomes for each student and teacher over time because the EMIS system provides IDs to follow individuals longitudinally if they remain in the system. The primary use of the EMIS data would be to estimate impacts on enrollment, because estimates of impacts on test scores would be biased (as we explain in the following paragraphs).

The larger sample sizes provided by administrative data, compared to survey data, will improve on the MDEs for all outcomes at the student, teacher, and school levels. Table III.5 provides the MDEs for estimating project impacts at the student, teacher, and school levels by using this data at endline. We assume in our calculations that we will be able to use a full sample of 261 LS schools and 150 US schools in the analysis. Using administrative data will allow us to detect impacts as small as 0.18 and 0.25 standard deviations at the student level for LS school impacts and US school impacts, respectively, for outcomes such as enrollment or completion. It will also allow us to estimate impacts as small as 0.20 and 0.29 for teacher-level outcomes at LS and US schools, respectively. The MDEs for outcomes such as the enrollment or completion rates at the school level improve to 0.35 and 0.50 at the LS and US level, respectively.

Table III.5. MDEs for student-, teacher-, and school-level outcomes, using administrative data for the full sample of schools in random assignment

	Treatment schools (Number)	Control schools (Number)	MDEs for student-level outcomes (standard deviations)	MDEs for teacher-level outcomes (standard deviations)	MDEs for school-level outcomes (standard deviations)
LS school impact	57	204	0.18	0.20	0.35
US school impact	27	123	0.25	0.29	0.50

Note: MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. We based the total number of eligible schools on MCC program documents and data from the Tanger-Tétouan-Al Hoceima region, where data is currently available. We estimated the total number of schools by assuming that the number of eligible schools in the study areas similar to the number of eligible schools found in the Tanger-Tétouan-Al Hoceima region. We plan to collect administrative data for an average of 266 students per school in LS and 332 students per school in US. We base this assumption on the average student population per school for Morocco secondary schools, as provided by the Education Policy and Data Center (2015). We assume attrition to be 10 percent for students at endline. We assume that the proportion of individual-level variance in the outcome explained by covariates for students and teachers is 0.40 and that the proportion of group-level variance explained by covariates is 0.30. We apply an intra-cluster correlation (ICC) of 0.25 as estimated from the TIMSS test score data for 8th-grade Moroccans in mathematics. We assume the same ICC for teacher outcomes.

One caveat is that the use of administrative data may produce biased estimates of impact, in particular for student learning. Students that drop out of school will not have test scores available at follow-up. If lower-ability students in treatment schools (who would have otherwise dropped out) but are encouraged to stay because of the MIAES lower a school's average test score, then we might erroneously find that the interventions decreased student learning in these schools relative to control schools, which had many of these students drop out. Despite this limitation, we believe that collecting administrative data on student test scores is still valuable. We can address the bias arising from the use of such data for test scores in two ways. First, we can use estimated impacts on enrollment rates to place upper and lower bounds on the estimated impacts on student-level outcomes (Lee 2009). Second, we can attempt to control for student characteristics that capture compositional differences and might be associated with the outcomes of interest. The most relevant characteristics would be measures of each students' pre-intervention academic achievement. The EMIS administrative data will enable us to assess the extent to which student dropout is an issue for the analysis. If it is not, then we can verify the use of administrative test scores to estimate impacts by comparing with impacts we estimate using survey data (which will not suffer from the same problem).

c. MDEs for pooled lower and upper secondary data, using survey and administrative data

Although MCC is primarily interested in estimating impacts at each of the LS and US school levels, we will explore estimating pooled impacts using administrative and survey data whenever there are common outcomes recorded for both levels. This analysis will combine data from both school levels. The advantage of this approach is that it improves on statistical power by utilizing all available data to estimate impacts. We will interpret the impact estimates as the average impact of the MIAES in LS and US schools. Table III.6 shows the MDEs from this approach; the estimated MDEs are smaller (better) than for the other analyses.

Table III.6. MDEs for pooled LS and US school impacts

Data Source	Treatment schools (number)	Control schools (number)	MDEs for student-level outcomes (standard deviations)	MDEs for teacher-level outcomes (standard deviations)	MDEs for school-level outcomes (standard deviations)
Administrative data	84	327	0.14	0.16	0.29
Survey data	84	84	0.20	0.21	0.36

Note: MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. We use the same assumptions from previous tables but pool data on LS and US schools in the analysis to estimate MDEs.

4. Data collection

Our RCT approach calls for two rounds of quantitative data collection: the first at baseline and the second three years later at endline. The survey instruments will include a longitudinal student survey, a student skills assessment, a teacher survey, direct classroom observation of teachers, a school director survey, and a school infrastructure checklist. These data are in addition to the administrative EMIS data that we will obtain from the Ministry of Education.

Table III.7 presents an illustrative list of outcomes that each data source would collect during each round of data collection. With Mathematica's support and oversight, a local firm procured by MCA-M would collect these data. We will provide the final list of the outcomes that we will collect as more details of program implementation become available. We provide a description of each of the quantitative data sources below.

Longitudinal student survey (N = 2,520 students, 15 students per school). The longitudinal student survey will gather basic data on student demographic characteristics at baseline, which will be used, among other things, to confirm that random assignment succeeded in creating equivalent groups of treatment and control students. These data will also be used to control for any remaining imbalance in the groups that occurs by chance, to get at unbiased and more precise estimates of the impact of the program. In addition, we will obtain contact information for students in the sample to facilitate tracking them over time. At follow-up, the survey will gather information on recall-based measures of student attendance, enrollment, and other student outcomes.

Student skills assessment (N = 2,520 students, 15 students per school). Measuring whether activities lead to improvements in student learning is a crucial component of the evaluation. However, given that many students eventually drop out of secondary school, we cannot rely on exit exam test scores to evaluate student learning because many youth in the sample will not take these tests. Thus, we will assess students on numeracy, literacy, and soft-skills at follow-up using tests that will be administered right after the longitudinal student survey. Our assessments will measure student learning, even for the youths who drop out of school. Baseline will also incorporate assessments, but will measure a smaller set of outcomes, since baseline test scores from the MENFP will be available for the full sample.

Table III.7. Data sources for the RCT

Data source	Outcomes
Longitudinal student survey	Attendance and enrollment Student aspirations and goals
Student skills assessment	Numeracy and literacy Soft skills (such as critical thinking, self-esteem, self-control, perseverance and social skills)
Teacher survey	Knowledge of and attitudes toward pedagogical innovations Self-reported teacher practices such as attendance, hours of instruction, and use of technology
Classroom observation	Hours of instruction Time on task Student time spent studying Use of pedagogical innovations from teacher training
School director survey	Use of school budget Content and status of school improvement plans Description and composition of school management committees School registrar records of teacher and student attendance (if available)
School infrastructure checklist	Measures of overall infrastructure quality and physical environment of the school
Administrative EMIS data	Student enrollment, completion, and dropout Baseline student exam scores Teacher attendance[a]

[a]If it is possible to obtain EMIS data about teachers, we will use information about teacher attendance in the impact analysis. See Appendix A for additional information linking the research questions to outcomes.

In assessing numeracy and literacy, we envision our tests to include a combination of questions from the Moroccan secondary exit exams and international standardized instruments such as those used for the TIMSS. We will work with the Ministry of Education to develop the tests and ensure that they contain the appropriate content and are not burdensome to administer. We will adjust the difficulty of these tests to be at the level of knowledge that students are expected to achieve at their grade level.

The literature points to several approaches to measuring soft skills. One approach is to rely on self-reported answers to a range of survey questions. Different surveys in the United States opt for this approach and are available as templates (Wilson-Ahlstrom et al. 2014). These surveys ask students to rate themselves on categories such as, “I don’t blame others for my mistakes” or “I am actively engaged in learning new things.” In its Skills Towards Employability and Productivity Program (STEP), the World Bank has also developed a set of surveys that are designed to measure personality traits, grit, and behavior in multiple countries (World Bank 2014). Mathematica has experience fielding such surveys in other projects. One might be concerned however of the subjectivity of these measures if students are merely reporting socially acceptable answers.

An alternative approach is to conduct a direct observation of soft skills. For example, the study by Groh et al. (2015) used psychometric testing to assess the soft skills of young job seekers in Jordan. The assessment consisted of three interactive exercises in the form of a group exercise with peers, a role-playing exercise, and a skills-based interview. Soft skills specialists then rated the behavior of students in these activities in terms of listening, self-confidence,

initiative, organization, and other behaviors. This approach provides a more objective assessment, yet might be time-consuming to implement. Other concerns are assessor competency and lack of knowledge on how to properly assess soft skills.

We will use an internationally validated assessment that gathers data on the soft skills listed in Table III.7 at baseline. We will meet with stakeholders to discuss the measurement of soft skills for the endline once we know more about implementation of project activities. One challenge is that soft skills is a general concept and may refer to many traits in multiple domains such as critical thinking, self-esteem, self-control, social skills, or socio-emotional skills. The evaluation should ideally focus only on soft skills that the program intends to improve among students, because there is no reason to measure soft skills in domains in which we do not expect to see impacts. The implementation team has not yet finalized its plans on the soft skills it intends to target. We will fine-tune our approach once implementation plans are in place, and may use different outcome measures at baseline and endline.

Teacher survey (N = 1,008 teachers, 6 teachers per school). Apart from establishing baseline characteristics of teachers—their credentials and experience—the teacher survey will gather recall-based measures of attendance, hours of instruction, and use of technology in the current and previous school years at baseline and endline. We will ask teachers about their knowledge and use of certain pedagogical practices and their attitudes toward them. The instrument will be modified at endline to include additional measures as needed, depending upon the actual intervention activities that occur.

Classroom observation (N = 1,008 teachers, 6 teachers per school). We will conduct classroom observations to assess teaching practices and behavior change. These observations will serve as the primary source of information on teacher behavioral change. We will triangulate these data with the teacher self-reports in the survey to improve the validity and reliability of the data. The observations will record teachers' time on task, pedagogical practices, and use of technology to track improvements in instruction. We will conduct these observations at baseline and endline by using the same measurement approach, but will add measures at endline if needed.[18]

School director survey (N = 168 school directors of schools included in the survey sample). The school director survey will gather data on operations and maintenance practices of the school, in addition to information on the school budget and average expenditures at baseline and endline. We will also gather data from school administrative records on student enrollment, dropout, and absenteeism so that we can triangulate the information with what we will obtain via the EMIS data. Enumerators will also ask about intervention activities, such as the school's improvement plan and its progress on goals. These data will help in answering questions related to school autonomy and accountability (for the implementation study), as well as in assessing the size of the budget managed by schools and the budget's common uses.

[18] We plan to use the Stallings instrument (Bruns and Luque, 2014). The Stallings Observation instrument is an internationally standardized and validated instrument that collects classroom data on four main variables: teacher's use of instructional time; teacher's use of materials; teacher's core pedagogic practices; and teacher's ability to engage students in the learning process.

School infrastructure checklist (N = 168 schools included in the survey sample). During visits to schools, enumerators will visually assess the quality of the school's infrastructure at baseline and endline. We will train enumerators to inspect the condition of classrooms, the main building, toilet facilities, laboratory equipment, and corridors. The data will be used to determine whether the infrastructure component of the project resulted in a more conducive learning environment for students.[19]

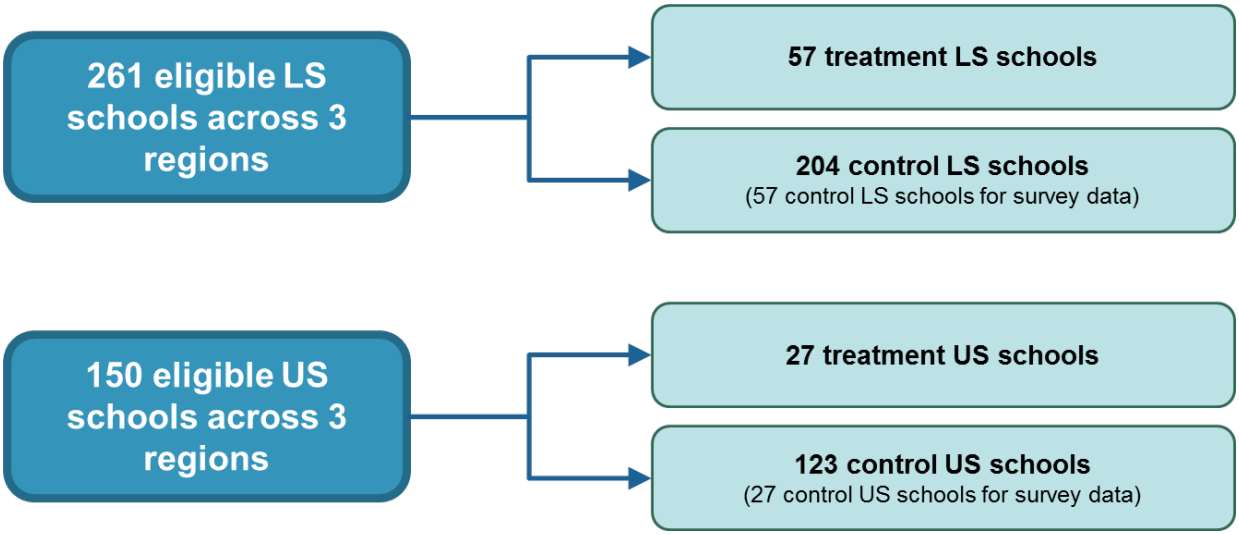
Administrative EMIS data. We will work with the Ministry of Education to obtain access to longitudinal, individual-level administrative records of students and teachers. These records include information on each student enrolled in each school by grade level in a school year. We are particularly interested in the enrollment status of each student and information on the student's latest grade level completed. We are also interested in national exit exam scores taken by students at the end of US school and regional exams taken by students at the end of primary and lower secondary school. We understand that data on exam scores will be easy to match with data on students through the student IDs used in the EMIS. We expect similar records to be available for teachers that will provide information on their subjects taught, credentials, and attendance. We would like to obtain these data from the Ministry of Education annually at the start of each school year (for data from the previous school year and enrollment from the current school year) for all students and teachers in all schools that took part in random assignment. Using these data, we will be able to estimate impacts on enrollment and attendance for a longitudinal sample of individuals more precisely than with the survey data, provided that these data are of usable quality.

5. Analytical approach

The approach to estimate impacts will involve comparing mean outcomes of the treatment and control groups at endline. Figure III.3 illustrates our approach. We will calculate the difference in mean outcomes between students, teachers, and schools in the 57 treatment schools and the corresponding groups in the 204 control schools for administrative data (or in the balanced sample of 57 schools for survey data) to estimate impacts at the LS school level. We will do the same to estimate impacts for the US school level by comparing the 27 treatment schools to the 123 control schools for administrative data (or to the balanced sample of 27 control schools with survey data). Any differences between the groups provide unbiased estimates of the effect of the MIAES as a result of having randomly assigned schools to each group. Because the student assessment and EMIS and the O&M subactivities are implemented in all secondary schools, the effects of these activities will balance out in both groups.

[19] We considered the option of hiring engineers to evaluate the state of the infrastructure in schools because they would be able to conduct more thorough and accurate assessments. However, after some discussion with MCC, we determined that this would not be a good use of resources. MCC explained that the government would have a separate quality assurance plan in place to determine that infrastructure improvements are carried out correctly. It would be sufficient for the evaluation to conduct a casual inspection of school surroundings and to ask teachers and students their perceptions of the school environment in order to understand the effect of infrastructure on learning.

Figure III.3. Comparison of treatment and control groups provides impact estimates



We intend to use regression analysis to improve on estimates provided by the simple comparison of means. Regression analysis improves on the statistical precision of estimates by incorporating key design features, such as stratification in random assignment. It also accounts for differences between treatment and control groups that may arise by chance. Hence, we will calculate impacts of the MIAES on key outcomes by estimating the parameters of the following equation using ordinary least squares:

$$(1)\quad Y_{ij} = \alpha + \beta T_{j} + \gamma X_{ij} + Z_{j} + \delta_{k} + \varepsilon_{ij},$$

where Y_{ij} is the outcome of interest for individual i in school j at endline; T_j is a binary variable that is equal to 1 for those in a treatment school or 0 for those in a control school; X_{ij} and Z_j are a set of individual and school characteristics that are measured at baseline (such as gender, age, test scores); δ_k is a vector of indicators that account for stratification in random assignment; and ε_{ij} is a random error term. The estimate of the coefficient, β, provides a regression-adjusted estimate of the impact of the MIAES. Equation (1) will be estimated independently for LS and US school impacts by restricting the sample to either of these groups. We will apply weights in our analysis to account for the different sampling probabilities for students and teachers and for nonresponses that may systematically occur for certain types of individuals.

Our estimates have to account for the fact that outcomes among individuals in the same school—the level of random assignment—are likely to be correlated because they experience many of the same conditions (such as the same teachers). We will account for the correlation statistically by clustering the regression error terms at the school level to adjust the standard errors, and will do so for student- and teacher-level estimates. For school-level estimates, no clustering adjustment will be necessary because all terms in Equation (1) will be at the school level.

Additional analyses. In addition to the analyses described above, we will conduct pooled analyses of US and LS schools, gender subgroup analyses, and robustness checks to incorporate weights. To estimate the pooled impacts of US and LS schools together, Equation (1) will be estimated by using all schools. The impacts for a particular subgroup can be evaluated simply by restricting the sample used to estimate Equation (1) or by including appropriate interaction terms in this equation. The subgroup analysis will be estimated separately for LS and US schools, as well as in a pooled analysis.

We will also explore conducting subgroup analyses by income level of school district, urbanicity, and distance of a student's residence to the school. While we will use the most rigorous methods to analyze this data, it is important to note that the results related to these disaggregation may be limited by statistical power since the sample sizes we propose do not explicitly guarantee that there will be enough data and variation on these subgroups to estimate effects.

Weights. We will also conduct robustness checks with alternative weighting schemes to ensure that findings are robust to alternative specifications of the model. In these checks, we will include school-level weights to account for slightly different probabilities of receiving the MIAES intervention and weights to account for different sampling probabilities for students and teachers because of varying numbers by school, in order to make the sample representative of students and teachers. We will include nonresponse weights to try to account for possible systematic attrition of certain types of students, if needed.

C. Qualitative study of the MIAES, assessment and EMIS, and O&M subactivities

In this section, we describe the data sources and analytic approach for the second component of the evaluation of the Morocco Secondary Education project—a qualitative study. This study will draw on two rounds of interviews with key stakeholders and one round of interviews and focus groups with teachers, school directors, students, and parents. The first round of interviews (2018) will capture information from ministry and regional education officials to understand the challenges they face in the secondary education system and their hopes and expectations of the project. The interviews will help us understand how assessment and EMIS data are currently used in Morocco. The second round of data collection (in 2020), which will include both interviews and focus groups, will capture information on how the project was implemented and about perceptions of the effects of the interventions.

The qualitative study will complement the impact evaluation in several ways. First, it will enable us to explore how, why, where, and for whom perceived and actual changes in outcomes occurred over time. Second, for questions related to areas that the impact evaluation does not target (that is, assessment and EMIS system changes, institutionalization of O&M), the qualitative study will be the primary source of information about changes in these outcomes. Finally, it will enable us to answer the research questions related to the sustainability of the program, which the other components of the evaluation will not be able to illuminate.

1. Design of the qualitative study

The qualitative study will document intervention plans, how these plans change over time, how they were implemented, and beneficiary perceptions of intervention effects. The results will help us understand project implementation, the facilitators of and barriers to change, and the accessibility and acceptability of the activities to the target population. The study will also provide early results (in 2020) to MCC, GoM, and the evaluation team so that we can understand any drivers of impact that will be measured by the impact evaluation endline. The study will triangulate information from both primary and secondary data sources, including interviews, focus groups, and program documentation. For example, the interview and focus group data will give meaning to any numerical changes that we see in student enrollment, retention, and completion. These data can also help us strengthen how we interpret any changes in student outcomes in the final impact evaluation report. The qualitative study will gather perspectives from multiple stakeholders: students, teachers, school directors, parents, and key informants from the Ministry of Education. We will systematically categorize and sort the qualitative information from these respondents to identify patterns and key themes to inform answers to the research questions. In addition, where appropriate, we will incorporate information from the quantitative data being collected for the impact evaluation, including the EMIS data and survey data.

2. Qualitative study sample

Our study sample for the qualitative study will be focused in the three regions receiving the MIAES and will include respondents from treatment and control schools. We will also conduct interviews at the national and regional level for all ten regions in the country to verify that the findings for those three regions are representative of the other regions in the country.

Follow-up sample. We propose a purposeful sample of teachers, students, and parents (including those who are members of the school management committee) from a subset of treatment and control schools in the RCT for data collection during the follow-up of the implementation study. We will use maximum variation sampling, in which participants are selected based on belonging to a high-, medium-, or low-performing school.[20] This sampling approach allows us to identify themes that occur consistently across schools and participants regardless of school performance, to identify issues that may be unique to any one group, and to closely examine schools that move between thresholds at follow-up to understand how the interventions contributed to the movement. We will use administrative data (including test performance) to identify thresholds for the high-, medium-, and low-performing schools and will randomly select 6 LS schools and 3 US schools per region, for a total of 27 schools. Table III.8 displays the resulting number of schools per group. We will make sure to include schools from both large and small schools and urban and rural areas.

[20] We will use national (for US schools) or regional (for LS schools) learning assessment results at the school level to determine the thresholds for high-, medium-, and low-performing schools.

Table III.8. School sample size for the qualitative study, by school type

	Low performing		Medium performing		High performing	
Regions	**LS**	**US**	**LS**	**US**	**LS**	**US**
Tanger-Tétouan-Al Hoceima	2	1	2	1	2	1
Fès-Meknès	2	1	2	1	2	1
Marrakech-Safi	2	1	2	1	2	1
Total number of schools	**6**	**3**	**6**	**3**	**6**	**3**

Note: We will use national learning assessment results at the school level to determine the thresholds for high-, medium-, and low-performing schools. Schools in each group will be randomly selected.

We will conduct focus group discussions with teachers, students, and parents, as well as in-depth interviews with school directors. The focus of the data collection effort is to understand the implementation process and how the interventions have changed participants' practices and improved the quality of the learning environment and education system. To minimize the costs of finding student and teacher samples, we will invite the same students (enrolled students and dropouts) and teachers interviewed for the RCT to participate in the focus groups. We will request that the data collection firm also invite to participate students from the longitudinal sample who have dropped out of the system. If they are unable to attend the focus group, we will consider visiting them in their homes for a short interview so that we can include their perspective on the program. Our ability to interview students at home will depend upon the time and cost of the qualitative study. We will invite the parents of students who are being tracked longitudinally to participate in the parent focus groups.

Nationwide interviews. We will supplement the discussions with students, teachers, parents, and school directors with key informant interviews with national, regional, and local Ministry of Education officials about details of the EMIS and O&M subactivities. We discuss further details in the next section.

3. Data Collection

The data collection effort will include in-depth qualitative data gathering through interviews with national- and district-level officials at baseline, a few open-ended survey questions for teachers and school directors at baseline, and focus groups and key informant interviews during a follow-up round. We will recommend bringing together stakeholders from all levels of the system for a meeting that discusses the facilitators of and barriers to change. The meeting would gather information from different groups of stakeholders. We would then work collaboratively to bring the various perspectives together to understand what aspects of the program are sustainable and which components may be experiencing less change. These data collection methods will be supplemented by a desk review of program documents.

We present a set of outcomes that we recommend measuring with each type of data collection in Table III.9. This list is based on our current understanding of intervention activities, and will potentially change as intervention implementation plans are finalized. We discuss each in more detail below. We will discuss these with project stakeholders to finalize the set of measures we will gather for the qualitative study.

Student focus groups (N = 27 focus groups). We will conduct focus group discussions with our student sample in 2020 to gather data on student perceptions of the Secondary Education Activity. The goal of these focus groups is to help us understand changes in pedagogical delivery, student perceptions of the interventions, and how the interventions have contributed to changes in the students' learning and schooling experience. We are particularly interested in their views of teacher quality and the school environment. To facilitate an environment where girls and boys would feel safe in expressing themselves, we will consider gender-disaggregated focus groups.

Teacher focus group (N = 27 focus groups). Similar to the student focus groups, we will conduct these group discussions in 2020. The focus groups will seek to understand what teaching practices teachers think they have learned, how they are using the practices in the classroom, what type of support they receive at the school, how parents engage with the school, and how they are using results of assessments to help improve learning for their students. We will triangulate the focus group results with findings from the teacher survey and classroom observations.

Table III.9. Qualitative data sources

Data source	Outcomes
Student focus group	• Perception of quality of instruction • How the interventions affected learning and schooling experience over time • How changes in the school environment affected a student's decision to stay in school (or drop out) • Types of changes in pedagogical delivery (student perspective) • How students have improved their foundational skills and soft skills • How students believe soft skills and technology will help them transition into the workforce • What the biggest changes in the quality of schooling are and why
Teacher focus group	• How teacher behavior has changed; what facilitated or hindered the changes • How (through examples) teachers assist students to learn and use soft skills • Teacher perceptions of students' abilities to demonstrate soft skills and improved foundational skills • Teacher perceptions of the contributions of the program • Perception of accountability among colleagues • How school improvement plans (SIPs) have been used at the school • The ways in which SIPs been successful or not at the school • Sustainability of SIPs • Changes in school leadership and management • How new student assessments and EMIS are used to help improve student learning
Parent focus group	• Knowledge of school improvement plans • Perception of changes in school management and accountability among teachers • Parent/community involvement in school activities • Parental perspectives on changes in student behavior, goals, and ambitions • Parental expectations of a student's future • Parental knowledge of a student's school performance (that is, grades, assessment results)

Table III.9 *(continued)*

Data source	Outcomes
School director in-depth interview	• Change in school management practices • Perception of autonomy and accountability of school • Parental/community involvement in the school • Changes in the school environment and its contribution to student outcomes • Perception of changes to teacher pedagogical delivery, support, and ability to teach soft skills • Perceptions of school-based management • How data is used to inform school needs and performance (that is, budgeting, assessment, accountability) • Perception of student behavior, including changes over time • Perceptions of future for students • Facilitators and barriers of change
Key informant interviews with officials at the Ministry of Education	• Perceptions of how the reforms have been implemented in the three regions • Strengths and weaknesses of the three subactivities • Sustainability of interventions • Facilitators of and barriers to change • Examples of systems or processes that have changed; how and why • Implementation of EMIS and student assessments • How data are being used to inform policy and decision making • Strengths and weaknesses of new O&M plan • Implementation process for O&M • Facilitators of and barriers to implementation • Sustainability of O&M
Stakeholder meeting	• Facilitators of and barriers to organizational change
Document review	• Implementation plans (including the Social and Gender Integration Plan) • Implementer quarterly and annual reports • Research studies conducted by implementers • Policy documents

Parent focus group (N = 27 focus groups). We will invite a subset of parents (including those who are members of the school management committees) linked to the sample students for a focus group discussion in 2020. We will explore their role in school management, levels of parent and community engagement in school activities, and how their role has changed over time. The discussion will explore the process for developing and implementing school improvement plans and its strengths and weaknesses.

School director in-depth interview (N = 27 school directors of schools in the focus group samples). These school director in-depth interviews, which will occur at the same time as the focus groups with teachers and students in 2020, will gather information related to changes in management practices and perceptions of autonomy and accountability among schools. These interviews will facilitate our understanding of how the school improvement plans translated into changes in schools.

Key informant interviews with national, regional, and local Ministry of Education officials, including MCC staff and implementers (N = approximately 19 individuals). We intend to conduct interviews with staff who are involved firsthand with implementing the MIAES, student assessment and EMIS, and O&M subactivities. These individuals include the minister of education; the heads of the AREF from all 10 of the northern regions; the education leads from MCA-M (2 individuals) and MCC (one individual); the implementer of the MIAES; the consultant who will design the improvement of the EMIS system and the counterpart official from the Ministry of Education; and the consultant who designed the national student assessments and the counterpart official from the Ministry of Education. We include these stakeholders because several of the activities extend beyond the three regions and will reach a national scope. The interviews will focus on understanding their roles vis-à-vis the reforms, their perspectives on any change that is occurring, and the facilitators of and barriers to success. These interviews will be prearranged and will occur during the qualitative data collection effort in 2020 for the implementation study. We will use the interviews from the seven AREFs that are not a part of the MIAES activity to confirm that the experience in the other regions with the EMIS and the O&M subactivities are consistent across regions (and are not fundamentally different for the three regions that receive the MIAES as well). Our discussions with officials throughout the course of the evaluation will also be considered a data source for the study.

Stakeholder meeting. We will conduct a stakeholder meeting to gather follow-up data on institutional change. The goal of the meeting is to identify examples of national- and regional-level change in education policies and practices; what facilitated the changes; and where change has not happened, understand why. The meeting will be held over a three-day period and include stakeholders from all levels of the system. Participants will be purposefully selected based on their roles in policymaking, budgeting, or planning (that is, ministry-level executive staff); in the implementation of program activities and new policies and practices (that is, implementers, regional education staff, and school directors); or as beneficiaries of the activities (teachers). The facilitator will guide stakeholders through identifying institutional and technical changes that occurred over the course of the project. The stakeholders help map the changes and bottlenecks in the system, bringing together the perspectives from various levels of the system. This process identifies the remaining barriers to institutional change and helps ministry- and district-level personnel identify areas that may affect sustainability of the program. The data gathered during this meeting may contribute to ministry planning processes. These data will also be triangulated with stakeholder interviews and focus groups for the final implementation study report.

Document review. Program document review will focus primarily on reports from implementers of the project activities (including the Social and Gender Integration plan), the O&M plans, policy documents, and relevant school records. This review will provide information on the implementation process, the main barriers to and facilitators of implementation, and key lessons learned. The document review could also identify topics of possible importance to explore in more detail through the qualitative data collection—for example, the qualitative work could explore particular ongoing challenges to successful implementation. We will develop a systematic review protocol to gather data from all relevant documents. We will review documents throughout the life of the evaluation, and will incorporate a larger set of documents to review at the time of follow-up for the implementation study.

4. Analytical approach

The qualitative study will explore (1) how teachers have changed their pedagogical practices, (2) perceptions of how the project has helped students prepare for the labor force, (3) how schools have used the school improvement plans, and (4) how institutions have changed or adapted to sustain the project interventions. The analysis will also explore institutional and technical practices that facilitated change and identify the barriers to change that remain in the system. We will complement the qualitative data with quantitative data from the impact evaluation.

Qualitative data analysis. The qualitative data analysis will explore how the components of the Secondary Education Activity integrate to improve accountability, provide an improved learning environment for students, and create a sustainable performance-driven education system. (Where appropriate, Mathematica will triangulate the quantitative and qualitative findings to ensure we provide depth and understanding to the analysis of each of the research questions.) This qualitative analysis will provide context and meaning to the impact evaluation findings and will help end users understand the roles of the different activities in improving the key outcomes. The qualitative study will include the following analyses:[21]

- **Examine the extent to which teachers changed their pedagogical delivery and why (or why not).** We will use interview transcripts to analyze examples of how teachers have changed their delivery to better teach students foundational and soft skills. We will triangulate this data with the classroom observation data collected in the impact evaluation to document whether we see the actual changes in the classroom.
- **Examine the extent to which students gained skills that better transfer to the workforce.** We will analyze student and teacher focus group transcripts to identify themes and examples of the skills that students perceive they have acquired and how they would use those skills in future employment. We will triangulate student and teacher responses and use the results to provide context and a deeper understanding to the skills assessment administered in the impact evaluation.
- **Examine the reasons that students chose to remain in school.** We will use focus group results to understand the decision making process that students use when deciding to remain in school or drop out. We will include current students and students who have dropped out of the program. For the students who have remained in school, the results will help us understand the extent to which project modifications helped that process. For those who dropped out of school, the focus groups will help us identify the reasons they left and what they are doing in terms of employment.
- **Examine how changes to the EMIS and assessment systems have addressed issues related to the use of data for decision making.** We will bring together the results of interviews with ministry officials, district education officials, school directors, and school-level focus groups with teachers to understand how changes to the EMIS and the assessment systems have facilitated or hindered the use of data for decision making. We will try to gain

[21] Please note that this is not an exhaustive list of the analysis we will conduct, but serves to provide examples of some of the main components of our analysis plan.

a deeper understanding of how these stakeholders use data and provide examples of how using data has led to policy changes or improved student performance.

- **Examine the extent to which school maintenance improvements and O&M led to an improved and sustainable learning environment.** We will use the results of interviews with Ministry of Education personnel, school directors, teachers, students and communities to examine the facilitators and barriers to sustainability. We will review these together during a stakeholder meeting, which will help us understand which improvements are likely to be sustainable and why.

We will follow four steps to analyze the qualitative data (Creswell 2009):

1. **Raw data management.** Raw data management is the process of organizing data into meaningful units of analysis (that is, from audio files to transcripts). During this step, we will review all data and eliminate any data that are incomplete or not useful to our analysis.
2. **"Chunking" and initial coding.** Often referred to as data reduction, this step will allow us to read through the transcripts several times and obtain a holistic sense of the data. We will develop a detailed initial coding scheme. We will map the coding scheme to the research questions and logic model. We will also develop internal summaries of results, trends, and patterns in the data to accompany the broader coding themes.
3. **Detailed coding.** This step will involve refining the coding scheme and recoding data as we look at the data in greater depth. We will use NVivo software to review and code the transcripts based on the initial codes developed during the chunking process. Using NVivo to assign codes to the qualitative data will enable us to access data on a particular topic quickly and organize information in different ways to identify themes and compile evidence supporting them. We will expand and refine these codes during the coding exercise and subsequent analysis of the coded transcripts in an iterative process as additional themes emerge. Further, the software allows respondents to be categorized by gender, age, geographic location, or other salient characteristics to permit analysis by group.
4. **Data interpretation and writing.** The analysis of the coded transcripts will involve triangulating the findings across stakeholders to highlight mechanisms, context, and similarities and differences in perspectives. The baseline and final reports will use the qualitative data to explore the implementation and results of the program activities fully.

Analysis of institutional change and sustainability. To understand how institutions are changing and the extent to which the MCC program is sustainable, we will use several analysis techniques as a lens for reviewing the qualitative data related to sustainability. The goal of the analysis will be to understand whether innovations introduced under the program are institutionalized over time and whether MIAES can be taken to scale nationally. This involves analyzing the facilitators of and barriers to change related to the program. We will aim to establish what, how, and why changes are happening in the education sector as a result of the three subactivities, as well as to identify any key bottlenecks in the system that may prevent the project from reaching its outcomes. We will use the data we collect through interviews with key stakeholders and document review. Further, we will conduct a stakeholder meeting at follow-up that allows us to bring together various stakeholders into a group setting and discuss the facilitators of and barriers to change. The process will allow the Ministry of Education to

consider action steps to move toward institutionalizing aspects of the reform. In addition the analysis will also shed light on facilitators and barriers to expanding the MIAES interventions to more schools (nationally).

D. Ensuring data quality

Mathematica, in partnership with MCA-M and MCC, is committed to ensuring that the data collected for the impact evaluation meet the highest data quality standards so that the results used for advising policy are precise and reliable. There are several steps that we will undertake to ensure that data quality is maintained, including the following:

- Work with MCC and MCA-M to hire the data collection firm. We will work with MCC and MCA-M to draft the terms of reference (ToR) for hiring a data collection firm. We will help MCA-M review the firm proposals and select the best firm based on the evaluation criteria set by MCA-M.
- Adapt existing survey instruments. We propose to draw on existing surveys developed for the impact evaluations of MCC's school investments in Burkina Faso, Niger, and Georgia. These projects are conducting similar interventions, including infrastructure improvements, community engagement, teacher training, and management training for principals. By adapting the existing surveys we can save time and resources, while using instruments that have been proven in the field. We will also incorporate validated instruments used elsewhere where relevant (for instance, we will use the Stallings instrument (Bruns and Luque 2014) to conduct classroom observations).
- Pilot the data collection instruments. We will work with the data collection firm to conduct extensive pilot testing of all data collection instruments in French and Arabic (and possibly Berber, if applicable) to identify any potential issues with the comprehension, flow, or cultural appropriateness of the instruments. We recommend that the pilot test take place in at least three schools. The pilot test process includes training enumerators, piloting data collection, documenting any needed instrument changes, and providing cleaned data sets and instruments. All key staff from the data collection firm participate in the pilot.
- Participate in data collection training. We will support the data collection firm to ensure that the enumerator training workshop is comprehensive and includes an in-depth explanation of the questions on each instrument, highly detailed protocols, and practice exercises for training assessors and classroom observers. The enumerators will have an opportunity to practice using the instruments in the field during the training process. Training participants will be required to attend all sessions of the workshop and demonstrate their acquisition of appropriate skills through supervised practices, exercises, or tests.
- Provide guidance on data collection protocols. We will provide guidance to MCC and MCA-M so that the data collection firm follows strict data collection protocols articulated in a well-written manual. We will further develop detailed manuals for data entry and cleaning to reduce errors stemming from these processes. The manuals will include explanations of all survey questions; data collection protocols and procedures (for example, consent, guidelines for protection of human subjects, how to approach a respondent, building rapport, and follow-up procedures if schools require revisiting); and clear guidance on the administration of the student assessment and classroom observation.

- Provide data quality assurance. The data collection firm ToR will require that data collection supervisors review each instrument and interview transcript immediately following the data collection process. Supervisors will require data collectors to return to respondents if they skipped any questions or if responses are ambiguous. Supervisors will also observe each assessor and observer during his or her first administration of the student assessment or classroom observation. They will retrain field staff or otherwise ameliorate difficulties if systematic problems are found. Mathematica and MCA-M staff will also conduct quality assurance on the data collection process by observing interviews, assessments, and classroom observations.
- **Test data entry system.** Mathematica has experience using both electronic and paper-based data collection systems. If data collection is conducted on paper, the data collection firm will develop the data entry system and provide a protocol for data entry and cleaning to Mathematica for approval. The data collection firm will test the data entry system by entering pilot data and fix any problems that are identified. The data collection firm will manage double data entry of all instruments, run frequencies on all variables, and provide this information to Mathematica in electronic form. The data collection firm will send the data electronically to Mathematica for data checks after the first 5 percent of cases are entered. If data is collected electronically, the data collection firm should already possess the necessary hardware capabilities for conducting electronic data collection. The data collection firm will test the electronic data collection system and fix any problems that are identified in the testing process. The data collection firm will transmit the data electronically to Mathematica for review as soon as it has been collected in the field and reviewed by the firm's supervisors. Mathematica will also conduct random audits of a sample of instruments to ensure that the data collected and entered are reliable and accurate.

E. Cost analyses

To assess whether its investments are sound, MCC uses economic rate of return (ERR) models to calculate the cost-effectiveness of its projects. The ERR is a summary statistic that captures the overall merits of an investment. Conceptually, it is the discount rate at which the project's benefits equal its costs. The higher the ERR, the greater the benefits of the project relative to its costs. Prior to compact signing, MCC completed its initial ERR analysis of the Secondary Education Activity. The ERR model defined the beneficiaries as all graduates of activity-supported secondary schools and their families. MCC calculated an ERR in the range of 12.4 percent to 15.1 percent for these households for a period of 20 years. MCC is in the process of updating these estimates and our evaluation report will use these estimates to calculate the final project ERR.

The ERR model is based on several key parameters and assumptions. MCC identifies three benefit streams that will support its investment. First, improvements in classroom pedagogy will raise student test scores and result in higher future earnings. Second, the implementation in school-based management will also increase student test scores and increase future earnings. Third, infrastructure improvements will lower dropout rates in school and increase completion. Using parameters from the literature, the model assumes effect sizes that would result from the interventions and estimates the expected number of students to benefit from these activities by referring to a household survey. The key parameters include the estimated effect of each of the

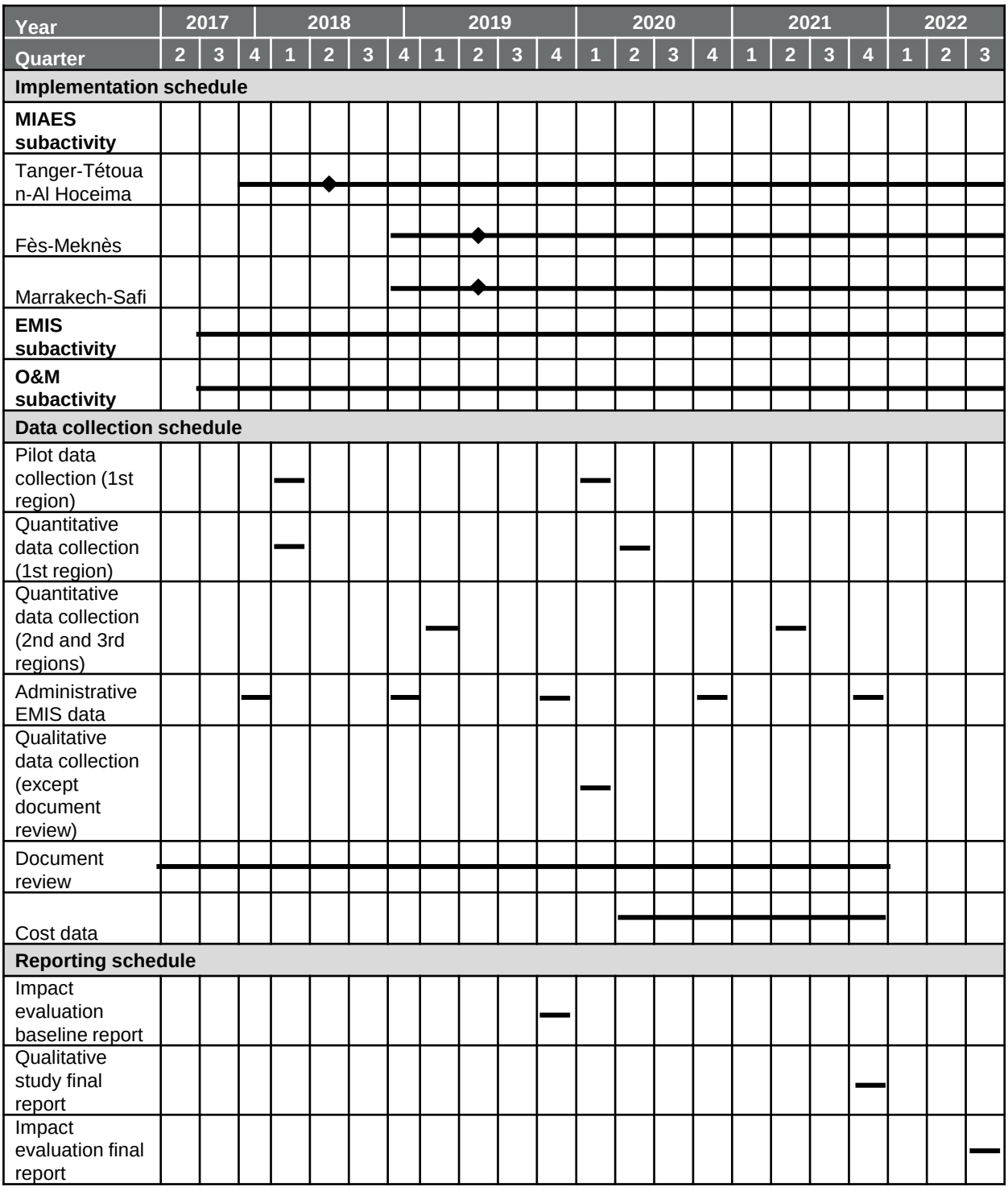
interventions on school transition, dropout, completion, student test scores, employment, and earnings.

As part of our evaluation, we will update the ERR model based on parameters estimated from our evaluation. Although it will not be possible to measure long-term impacts on employment and income within the time frame of the evaluation, we will be able to measure impacts on school attainment and learning as measured by test scores. We will use these estimates in calculating an ex-post ERR to the project to capture its cost-effectiveness.

F. Overall timing of implementation, data collection, and reporting

The timing of the evaluation activities that we have proposed is based on our understanding of the rollout of the interventions in schools and nationwide, our ability to measure unbiased impacts in MIAES activities, and our ability to provide information through the qualitative study that could be used by stakeholders in discussions at the end of the compact. We summarize our proposed timing for data collection in Figure III.4. The timing may shift depending upon actual implementation. Quantitative data collection will occur at baseline in the first year that schools receive the MIAES and three years later at follow-up (in 2020 for Tanger-Tétouan-Al Hoceima and in 2021 for Fès-Meknès and for Marrakech-Safi). The intention is to measure impacts for students who would have been exposed to the program for three years, during either LS or US schooling. This will enable us to estimate unbiased impacts for the entering cohorts of students in each school level after the maximum number of years of exposure at that level. In addition to the quantitative survey data, we recommend using national EMIS data, which would be requested from the Moroccan Ministry of Education. At the beginning of each school year, we would like to obtain start-of-the-year data as well as end-of-the-year data from the previous school year. We recommend that the endline qualitative data collection occur in 2021 in all regions so that findings can be produced and disseminated prior to the compact's closeout, and therefore included in GoM discussions regarding scale-up. Document review will occur throughout the course of the evaluation. Staggering the qualitative and quantitative follow-up will allow us to provide early results on the process of implementation to MCC, MCA-M and the MENFP. It will also allow us to adapt or focus the quantitative surveys to gather additional data (if needed) to support the impact evaluation.

Figure III.4. Implementation, data collection and reporting timeline



Note: A pilot program of the MIAES began in six schools in Tanger-Tétouan-Al Hoceima during the 2016–2017 school year and is not shown here. The black line indicates the timeline of implementation and data collection. The ◆ indicates the expected timing of implementation of the school infrastructure component for the MIAES.

G. Reporting schedule and dissemination plan

Mathematica will present baseline and endline evaluation findings in person to MCC and to stakeholders in Morocco. The timing of the analysis and reporting for the study will be determined by the program's phased rollout schedule (Figure III.4). Thus, Mathematica will present baseline findings after the end of the 2018–2019 school year, once baseline data collection has been completed in all three regions. We will present the qualitative study results at the end of 2021 and endline evaluation results toward the end of 2022. We plan 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.

- Mathematica will present a baseline report to MCC and relevant stakeholders once data collection has been completed in all three regions, after the end of the 2018–2019 school year. The baseline analysis will confirm whether or not random assignment successfully created equivalent treatment and control groups, while also describing the baseline characteristics of the study population.
- Mathematica will present the results of the qualitative study to MCC and relevant stakeholders in the fourth quarter of 2021, before the compact ends. The results will provide MCC, MCA-M, and the Ministry of Education with some early results on the implementation process, early successes, and lessons learned. The initial results will allow the Ministry of Education to begin planning for any scale-up activities.
- Mathematica will present evaluation findings to MCC and relevant stakeholders in the third quarter of 2022, after endline data collection and analysis have been completed in all three study regions for the RCT.

We will work with MCC to increase the visibility of the study's findings, particularly among education policymakers and development practitioners. We will collaborate with MCC and stakeholders to identify a variety of forums—including, conferences, workshops, publications, and online blogs—at various intervals to engage a wider audience while sharing results and encouraging donors, implementers, and policymakers to integrate the findings into future programming. For example, in addition to the project's full impact report, we will develop issue briefs summarizing and visualizing key findings from the final impact report for a broader audience of readers and stakeholders. Potential conferences for presenting evaluation findings will include forums hosted by the Comparative International Education Society, the American Evaluation Association, or the Association for Public Policy Analysis and Management. We will also seek to publish a peer-reviewed article disseminating the study's results in academic or sector-specific journals that are focused on education systems in developing countries.

H. Key challenges to the evaluation and strategies to mitigate them

We anticipate several risks to the evaluation, which will require careful monitoring and management throughout the study period. Below we list the key challenges to the evaluation and strategies to mitigate those challenges to the extent possible.

- **Introduction of other school-based programs.** The validity of our impact estimates relies on the assumption that, because of random assignment, the treatment and control schools will be the same in every way, except for implementation of the MIAES. This assumption is violated if during the study period other school level interventions are implemented in the treatment or control schools. If additional interventions were carried out in treatment schools only, then the interpretation of our impact estimates would change to be the impact of the MIAES plus additional interventions. The more concerning scenario is if the government or other donors provide more support to control schools to compensate schools that are not selected to receive the MIAES interventions. This additional support would invalidate our estimates of impact because the control group would no longer serve as a valid counterfactual to the treatment group. One example of a potential threat to the evaluation would be if the Workforce Development Activity were to target either treatment or control school areas. It is unclear at this moment which study schools might be involved in these projects. Another example would be if the GoM decided to improve the infrastructure in control schools but not in MIAES schools. We will work closely with GoM and MCC to prevent situations that would invalidate our impact estimates. It will be important for GoM to apprise the evaluation team of any plans for new activities that may not be rolled out to all schools equally. By knowing about such plans ahead of time, we can work with GoM to mitigate biased estimates of the MIAES.
- **Low statistical power for estimating impacts.** We anticipate that the impact evaluation might be underpowered to statistically detect small impacts for activities delivered at the US level, or for school-level impacts in general. Unfortunately, there is no easy way to improve statistical power in this case—unless the number of treatment US schools were increased, which is not possible because of program budget constraints. We will work to provide alternative evidence through the qualitative study to help MCC understand impacts that might occur at the US level or at the school level. Impacts observed at the LS level can serve as a benchmark for US school impacts. Through our implementation study, we will closely document ways in which program implementation differs between LS and US school contexts, and explore if it is reasonable to expect similar impacts at both school levels. In addition, we will be able to rely on our qualitative study to understand changes in school-level outcomes if we find that the effects we measure at the school level are not statistically significant but possibly substantial.
- **Tracking a longitudinal sample.** Tracking students over three years, when so many are expected to leave the education system, will be a challenge. Provided that students continue to live in our study areas, we will visit them at their residence to collect data at follow-up if we cannot survey them in school. We are concerned, however, that survey attrition rates could still be high. Attrition will limit our ability to estimate any statistically significant effects from the MIAES for student outcomes. It may also bias our impact estimates if attrition occurs disproportionately in either the treatment or control group. To mitigate this, we will devote extra resources to collecting high quality contact information for our study sample at baseline to help us locate the student sample after three years. We will collect information on the students' current addresses, telephone numbers, and social media accounts. We will also collect contact information for their parents and relatives.
- **Implementation delays.** We understand that the consultant is still finalizing the implementation plans and that the interventions are being tested in the pilot schools, which

may result in a delay of implementation activities of MIAES in the first region. We will continue to assume the current implementation timeline for the evaluation and make contingencies for delays that may arise over the next few years. If implementation is delayed for the first region, we will consider moving baseline collection to the following school year for this region and consider studying the entering cohort of students in the following school year (rather than the entering cohort in the preceding year).

- **Incomplete rollout of activities.** We recognize that unforeseen circumstances may prevent the full rollout of program activities in all treatment schools. For example, it may be infeasible to carry out school infrastructure rehabilitation in some schools because design assessments may uncover serious structural flaws that the MCC budget cannot afford to fix. In this case, the set of schools assigned to treatment would not be the same as the set of schools that ultimately receives all MIAES activities. If implementation of all MIAES activities cannot take place in some schools due to unforeseen circumstances, we will nonetheless treat such schools as part of the treatment group in our analysis when comparing outcomes. Proceeding in this way would preserve the equivalence of treatment and control group schools that was established by random assignment. The interpretation of impact estimates (the intent-to-treat [ITT] estimates) would measure the impact of being offered the program. ITT impacts, although different from the actual impact of the program, are relevant because most interventions occur in real-life settings, in which planned interventions are not always delivered. We would also estimate treatment on the treated (ToT) estimates to better understand the impacts on those students who actually received the program.

- **Limited ability to isolate impact of specific MIAES interventions.** We will not be able to disentangle the impacts of specific MIAES interventions through the RCT because the interventions will be bundled in each school. We will be able to provide suggestive evidence on interventions that may have played a greater role in contributing to any impacts. We will rely on our qualitative implementation analysis, together with our focus group discussions with teachers and students, to understand how various interventions were implemented and how they interacted with each other to see which interventions worked better than others and may be responsible for driving results.

IV. EVALUATION ADMINISTRATION AND MANAGEMENT

A. Institutional review board requirements and clearances

Mathematica will prepare and submit an institutional review board (IRB) application for approval of the research and data collection plans. The application materials include three sets of documents: (1) a research protocol, which will draw heavily on the present design report and include more information about plans for protecting study participants' confidentiality and human rights; (2) copies of all data collection instruments; and (3) a completed IRB questionnaire that summarizes the key elements of the research protocol, plans for protecting participants' human rights, and possible threats to participants if their confidentiality were compromised. Based on prior experiences, we expect that the study will qualify for expedited review because it presents minimal risk to participants. If so, the IRB can typically review the application within one week of its submission.

IRB approval is valid for one year from the date of approval and must be renewed on an annual basis. We expect that the annual renewals will require minimal updates to the core application materials. In addition, if data collection instruments change substantially from those that the IRB approved, then we must reapply for approval. Small changes to the instruments (such as rewording or reordering of questions) do not require reapplication, but the finalized instruments must be submitted to the IRB for documentation.

After Mathematica drafts the IRB research protocol, we will coordinate with MCA-M to ensure the data collector and local stakeholders agree on the data collection protocol. Because Mathematica will not have a contractual relationship with the data collector, the data collector's contract with MCA-M must specify that it will abide by the IRB's recommendations. The data collector and Mathematica must also sign an IRB authorization agreement stating that the data collector will adhere to the IRB-approved data collection procedures and protocols.

B. Data access, privacy, and documentation

After producing each of the baseline and final reports, we will prepare corresponding de-identified data files and codebooks that can be made available to the public. These data files, user manuals, and codebooks will be de-identified according to the most recent guidelines set forth by MCC. The public use data files will be free of personal or geographic identifiers that would permit unassisted identification of individual respondents or their households. We will remove or adjust variables that introduce reasonable risks of deductive disclosure of the identity of individual participants. Mathematica will remove all individual identifiers, including names, addresses, telephone numbers, government-issued identification numbers, and any other similar variables. We will also remove unique and rare data by using local suppression, replacing these observations with missing values instead. If necessary, we will also use top and bottom coding, which would set upper and lower bounds to remove outliers and would collapse any variables that make an individual highly visible (because of geographic or other factors, such as ethnic classifications or languages spoken) into less easily identifiable categories. We will introduce random errors into any gathered geographic data (for example, global positioning system or geographic information system coordinates), which would displace urban points from zero to 2 kilometers, rural points from zero to 5 kilometers, and an additional 1 percent of rural points from zero to 10 kilometers. We would also introduce additional perturbation as deemed

necessary. Data perturbation will take place in a manner that will not significantly degrade the data.

C. Evaluation team roles and responsibilities

Mathematica's project team has extensive experience with conducting mixed-methods, multicomponent, large-scale evaluations in the field of education. **Mr. Matt Sloan** will serve as the program manager, acting as the primary point of contact for MCC. Mr. Sloan will manage the relationships with government agencies and other local entities and contractors, while supervising the evaluation design and implementation process and ensuring high data quality. **Dr. Emilie Bagby** is the senior analyst for education economics for this evaluation. Dr. Bagby will provide methodological leadership and technical oversight and will support the project team. **Dr. Audrey-Marie Moore** is serving as the senior analyst for teachers and teaching. Dr. Moore will lead the qualitative data collection and contribute to the quantitative data analysis of teacher behavior change and the qualitative analysis process. **Dr. Paolo Abarcar** is serving as a senior analyst. Dr. Abarcar will oversee the study's quantitative data collection and analyses. **Dr. Evan Borkum** is the senior advisor. Dr. Borkum will support Dr. Bagby in the design and implementation of the impact evaluation and will provide quality oversight to the work.

REFERENCES

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.

Bagby, Emilie, Kristine Bos, Anca Dumitrescu, Nicholas Ingwersen, and Matt Sloan. "Niger NECS Impact Evaluation Report." Washington, DC: Mathematica Policy Research, July 2017.

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.

Braun, Henry, and Anil Kanjee. "Using Assessment to Improve Education in Developing Nations." *Improving Education through Assessment, Innovation, and Evaluation.* 2006, pp. 1–46.

Bruce, Kerry, Fahmina Rahman, Catherine Villada. "Ex-Post Performance Evaluation of MCC's Education Activities in Ghana." Washington, DC: Social Impact, October 2017.

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

Bruns, Barbara, and Javier Luque. "Great teachers: How to raise student learning in Latin America and the Caribbean." World Bank Publications, 2014.

Burnett, Nicholas, and Shubha Jayaram. "Skills for Employability in Africa and Asia." Washington, DC: Results for Development Institute, October 2012.

Campuzano, Larissa, Randall Blair, Virginia Poggio, Ivonne Padilla, Seth Morgan, and Michael Brannan. "Evaluation of the Formal Technical Education Sub-Activity of the Human Development Project, El Salvador" Washington, DC: Mathematica Policy Research, June 2016.

Carr-Hill, Roy, Caine Rolleston, Tejendra Pherali, and Rebecca Schendel. "The Effects of School-Based Decision Making on Educational Outcomes in Low- and Middle-Income Countries." 3ie Systematic Review Report 21. London: International Initiative for Impact Evaluation (3ie), September 2014.

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.

Creswell, John. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Los Angeles, CA: Sage Publications, 2009.

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.

Darling-Hammond, Linda, Ruth Chung Wei, Alethea Andree, Nikole Richardson, and Stelios Orphanos. "Professional Learning in the Learning Profession." Washington, DC: National Staff Development Council, 2009.

Deming, David. "The Value of Soft Skills in the Labor Market." NBER Reporter 2017: 4

Dumitrescu, Anca, Dan Levy, Cara Orfield, and Matt Sloan. "Impact Evaluation of Niger's IMAGINE Program." Washington, DC: Mathematica Policy Research, September 2011.

Education Policy and Data Center. "Morocco." Washington, DC: Education Policy Data Center, 2015. Available at http://www.epdc.org/country/morocco. Accessed February 3, 2017.

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.

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.

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.

Groh, Matthew, David McKenzie, and Tara Vishwanath. "Reducing Information Asymmetries in the Youth Labor Market of Jordan with Psychometrics and Skill-Based Tests." *World Bank Economic Review,* vol. 29, supplement 1, 2015, pp. 106–117.

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.

Guison-Dowdy, Anna. "Skills for Employability: Economic Returns to Education and Skills." Washington, DC: Results for Development, 2012.

Hanushek, Eric A., and Ludger Woessmann. *The Knowledge Capital of Nations: Education and the Economics of Growth.* Cambridge, MA: MIT Press, 2015.

Hanushek, Eric A., and Ludger Woessmann. "The Role of School Improvement in Economic Development." National Bureau of Economic Research Working Paper No. 12832. Cambridge, MA: National Bureau of Economic Research, January 2007.

Heckman, James J., Jora Stixrud, and Sergio Urzua. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." Chicago, IL: University of Chicago, January 2006.

Heckman, James J., and Tim Kautz. "Hard Evidence on Soft Skills." *Labor Economics,* vol. 19, 2012, pp. 451–464.

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.

Joyce, Bruce, and Beverly Showers. "Improving Inservice Training: The Messages of Research." *Educational Leadership,* vol. 37, no. 5, 1980, pp. 379–385.

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.

Kirst, Michael, Trish Williams, and Edward Haertel. "Similar Students, Different Results: Why Do Some Schools Do Better? A Large-Scale Survey of California Elementary Schools Serving Low-Income Students." Mountain View, CA: EdSource, 2005.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. "The Challenge of Education and Learning in the Developing World." *Science,* vol. 340, no. 6130, 2013, pp. 297–300.

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.

Lee, David S. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies*, vol. 73, no. 3, 2009, pp. 1071-1102.

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

McEwan, Patrick. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Review of Educational Research,* vol. 85, vol. 3, 2015, pp. 353–394.

Millennium Challenge Corporation. "Morocco Education and Training for Employability Project Program Design Document Draft." Washington, DC: Millennium Challenge Corporation, May 2, 2016.

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

Moore, Audrey-Marie, Amber Gove, and Karen Tietjen. "Great Expectations: A Framework for Assessing and Understanding Key Factors Affecting Student Learning of Foundational Reading Skills." *New Directions for Child and Adolescent Development,* no. 155, 2017, pp. 13–30.

Muralidharan, Karthik, and Venkatesh Sundararaman. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *The Economic Journal,* vol. 120, no. 546, 2010, pp. F187–F203.

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

OECD/ Eric Hanushek / Ludgar Woessmann (2015), Universal Basic Skills: What Countries Stand to Gain, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264234833-en.

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.

Results for Development Institute. "Bringing Learning to Light: The Role of Citizen-Led Assessments in Shifting the Education Agenda." Washington, DC: Results for Development Institute, 2015.

Szekely, Miguel. "Towards Results-Based Social Policy Design and Implementation." Working paper No. 249. Washington, DC: Center for Global Development, April 2011.

Wasik, Barbara A., Shira K. Mattera, Crishana M. Lloyd, and Kimberly Boller. "Intervention Dosage in Early Childhood Care and Education: It's Complicated." Research Brief OPRE 2013-15. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2013.

Wilson-Ahlstrom, Alicia, Nicole Yohalem, David DuBois, Peter Ji, and Barbara Hillaker. "From Soft Skills to Hard Data: Measuring Youth Program Outcomes." The Forum for Youth Investment, January 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 Group. "Kingdom of Morocco: Promoting Youth Opportunities and Participation." Report No. 68731-MOR. Washington, DC: The World Bank Middle East and North Africa Region Sustainable Development Department, 2012.

World Bank Group. "STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills." Discussion Paper No. 1421. Washington, DC: Social Protection and Labor, July 2014.

World Bank Group. "World Development Indicators, 2017." Washington, DC: World Bank Publications, 2018. Available at http://data.worldbank.org/data-catalog/world-development-indicators. Accessed February 26, 2018.

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.

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

APPENDIX A
SUMMARY OF THE MOROCCAN SECONDARY EDUCATION ACTIVITY

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

Table A.1. Evaluation summary table for the Moroccan Secondary Education Activity

Research question	Method	Data sources	Outcomes
MIAES subactivity			
1. What are the impacts on learning (numeracy, literacy, soft skills)?	RCT	Administrative EMIS data and student assessments	Exit exam scores by literacy and numeracy; assessments in literacy, numeracy, and soft skills
2. What are the impacts on key educational outcomes, including enrollment, completion, and attendance?	RCT	Administrative EMIS data and longitudinal student survey	Enrollment, completion, and attendance Student aspirations and goals
3. Are there differential impacts by gender across student outcomes?	RCT	All of the above outcomes by gender	
4. What are the impacts on teaching in participating schools and how were they obtained?	RCT and qualitative study	Teacher survey and classroom observation	RCT: Knowledge, attitudes, and changes in pedagogical delivery or innovation; changes in the use of technology; changes in time on task Implementation: How and why
5. What are the impacts on teachers' attendance?	RCT	Administrative EMIS data, teacher survey, and administrative school records from school director survey	Attendance
6. Did the MIAES interventions improve school management and lead to improved accountability among teachers? If so, how?	Qualitative study	Student, teacher, and school director surveys; student, teacher, , and parent focus groups; school director interviews	Perception of quality of teaching and accountability among teachers
7. How have institutional autonomy and accountability manifested themselves in participating schools?	Qualitative study	School director survey and interviews	Presence and status of school improvement plans, presence of school management committees, and perception of school accountability
8. What are the impacts on the quality of the infrastructure and physical environment of the school?	RCT	School infrastructure checklist	Quality of main buildings, classrooms, toilet facilities, and corridors
9. How did the size of the budget managed by schools and the common uses of this budget change?	Qualitative study	School director interview	Amount and use of school budget by type of expenditure
10. How is the decentralization process being incorporated in schools?	Qualitative study	Teacher and school director surveys; teacher, and parent focus groups; school director interviews	Management practices of the school, perception of decentralization from teachers, and parent involvement in school activities and school improvement plans
Assessment and EMIS subactivity			
11. How did the interventions contribute to improved student assessments, data, and policy feedback in the EMIS system and lead to a more performance-driven education system?	Qualitative study	Key informant interviews with national, regional, and local Ministry of Education officials; teacher focus groups; school director interview	Main accomplishments of subactivity, barriers to success, and recommendations for future implementation

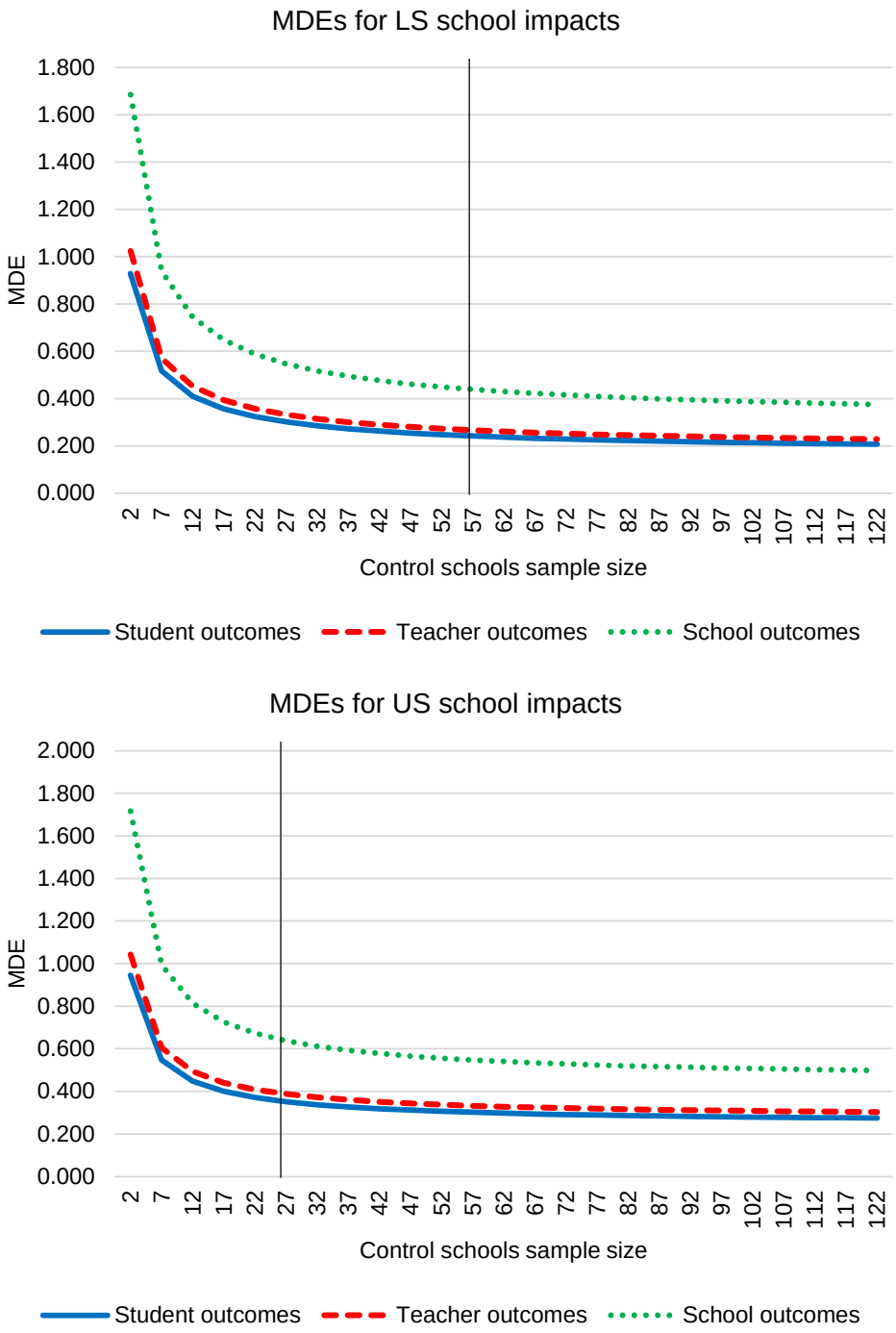
Table A.1. *(continued)*

Research question	Method	Data sources	Outcomes
O&M subactivity			
12. How do the infrastructure improvements and new O&M plan lead to an improved and sustainable learning environment?	Qualitative study	Key informant interviews with national, regional, and local Ministry of Education officials; teacher focus groups; school director survey	Main accomplishments of subactivity, barriers to success, and recommendations for future implementation
Overall sustainability			
13. To what extent can the Moroccan Ministry of Education sustain and scale the interventions under the Secondary Education Activity?	Qualitative study	Key informant interviews with national, regional, and local Ministry of Education officials; teacher focus groups; school director survey	Facilitators of and barriers to organizational change
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?).	ERR and Cost Effectiveness analysis Qualitative study	Use of quantitative data from the impact evaluation complemented with key informant interviews and focus group results as appropriate to inform the ERR findings.	Cost-effectiveness and financial sustainability of the Secondary Education Program.

APPENDIX B
STATISTICAL POWER FOR THE RCT

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

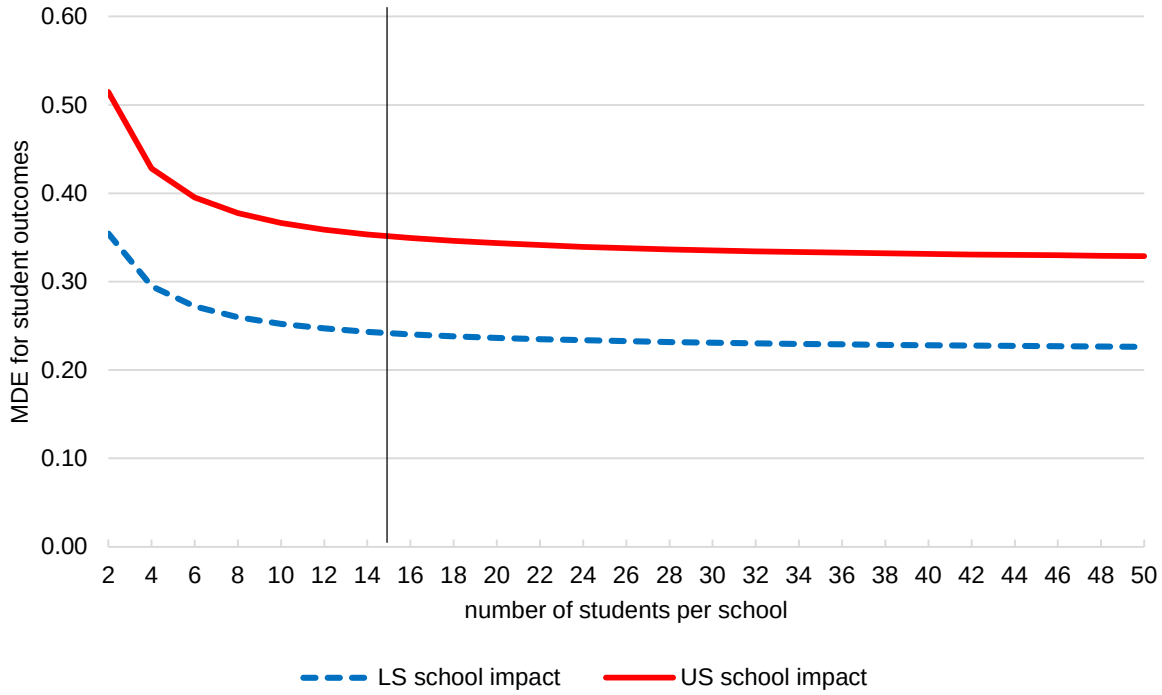
Figure B.1. MDEs for LS and US school impacts, varying the sample size of control schools



Note: The vertical lines mark the MDEs for the balanced sample of treatment and control groups for survey data. MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. We based the total number of eligible schools on MCC program documents and data from the Tanger-Tétouan-Al Hoceima region, where data is currently available. We estimated the total number of schools by assuming that the number of eligible schools in the study areas similar to the number of eligible schools found in the Tanger-Tétouan-Al Hoceima region. We assume data collection for all treatment schools in the sample. We hold constant the sample sizes for students and teachers at 15 students per school and six

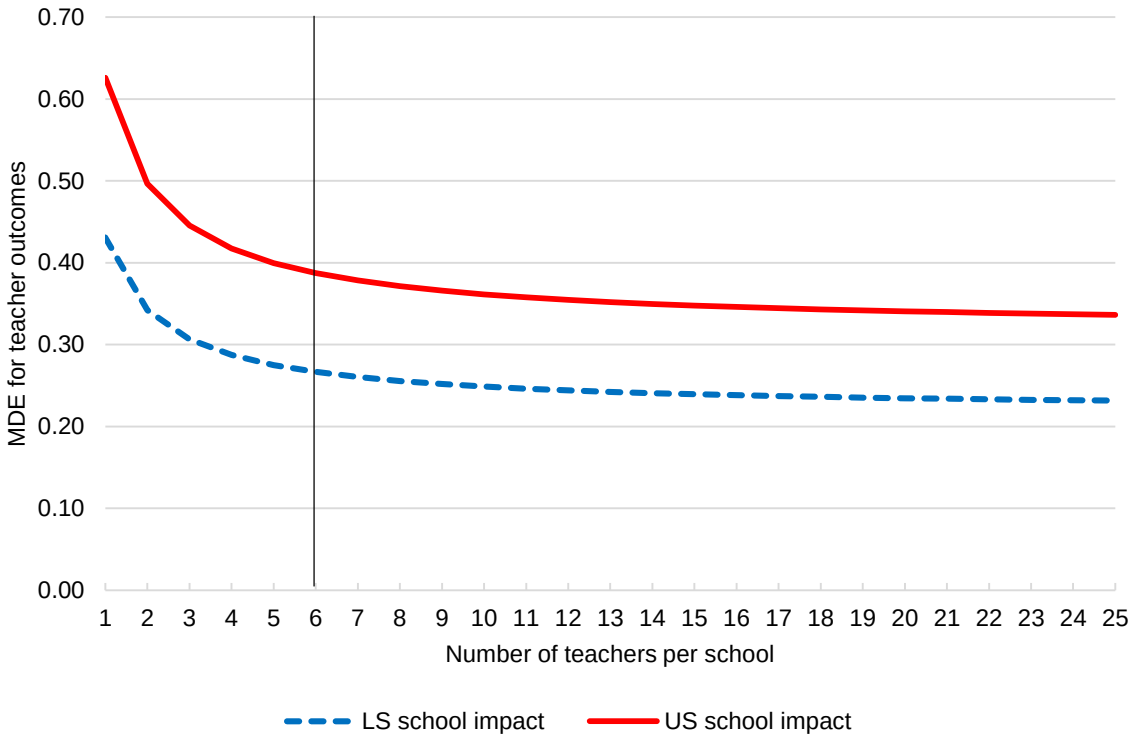
teachers per school. We assume attrition to be 20 percent for students and 10 percent for teachers at endline. We assume that the proportion of individual-level variance in the outcome explained by covariates for students and teachers is 0.60 and 0.15 and that the proportion of group-level variance explained by covariates is 0.30. We apply an intra-cluster correlation (ICC) of 0.25, as estimated from the TIMSS test score data for 8th-grade Moroccans in mathematics. We assume the same ICC for teacher outcomes.

Figure B.2. MDEs for student-level outcomes, varying the number of students per school to collect data from



Note: The vertical line marks the MDEs for our recommended sample size for students. MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. The calculations assu me a balanced sample of treatment and control schools to estimate LS and US school impacts. We expect attrition to be 20 percent for students at endline. We assume that the proportion of individual-level variance in the outcome explained by covariates is 0.40 and that the proportion of group-level variance explained by covariates is 0.30. We apply an intra-cluster correlation of 0.25, as estimated from the TIMSS test score data for 8th-grade Moroccans in mathematics.

Figure B.3. MDEs for teacher-level outcomes, varying the number of teachers per school to collect data from



Note: The vertical line marks the MDEs for our recommended sample sizes for teachers. MDE calculations assume a two-tailed test with a 95 percent confidence level and 80 percent power. The calculations assume a balanced sample of treatment and control schools to estimate LS and US school impacts. We expect attrition to be 10 percent for teachers at endline. We assume that the proportion of individual-level variance in the outcome explained by covariates is 0.40 and that the proportion of group-level variance explained by covariates is 0.30. We apply an intra-cluster correlation (ICC) of 0.25 (while it was estimated from the TIMSS test score data for 8th-grade Moroccans in mathematics, we also apply this ICC for teacher outcomes).

This page has been left for double-sided printing.

APPENDIX C
CALCULATING THE NUMBER OF SCHOOLS TO SELECT FOR TREATMENT IN EACH STRATUM

This page has been left for double-sided printing.

The procedure to calculate the number of treatment schools to select in each stratum can be represented with the following formula:

$$n_{t,u,p} = \left(\#of\ schools\ in\ strata\right) * \frac{x_{t.u}}{y_{t,u}},$$

where $n_{t,u,p}$ represents the number of schools to be selected in each stratum, t represents the school type (whether LS or US school), u is a binary indicator that represents whether the school is in an urban or rural setting, and p represents the province. The basic idea is to multiply the number of schools in each stratum by a weight, derived from the number of type t schools that are intended to be selected in an urban or rural area u ($x_{t,u}$), divided by the total number of schools of the same type in the same urban or rural area ($y_{t,u}$). Each school gets a roughly equal chance of selection for the treatment group. We round $n_{t,u,p}$ to be a whole number so that it provides the exact number of schools to select in a stratum.

As an example, in Tanger-Tétouan-Al Hoceima, MCC determined that it would implement the MIAES in a total of 28 schools (in addition to the 6 pilot schools)—19 schools at the LS level and 9 schools at the US level. Based on the geographic distribution of schools between urban and rural areas and across provinces, MCC determined that it would be appropriate to select 14 urban and 5 rural LS schools. Using the above formula, we chose one urban-LS school in Chefchaouen, 3 urban-LS schools in Larache, 6 urban-LS schools in Tanger, and 4 urban-LS schools in Tétouan. This selection process gave schools in each stratum an approximately equal chance of selection for the project.

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

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

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

SEATTLE, WA ■ PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA Policy Research