# Evaluation of the Niger Education and Community Strengthening Program

## Design Report

First Draft: February 14, 2013

Revised: August 23, 2013

Errata Corrected: October 1, 2014

Revised: June 6, 2016

Emilie Bagby
Evan Borkum
Anca Dumitrescu
Matt Sloan

MATHEMATICA
Policy Research

Mathematica Reference Number:
40038.300

Submitted to:
Millennium Challenge Corporation
1099 14th Street NW
Suite 700Washington, DC 20005-3550

Project Officer: Carolyn Perrin

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

# Evaluation of the Niger Education and Community Strengthening Program

## Design Report

First Draft: February 14, 2013

Revised: August 23, 2013

Errata Corrected: October 1, 2014

Revised: June 6, 2016

Emilie Bagby
Evan Borkum
Anca Dumitrescu
Matt Sloan

MATHEMATICA
Policy Research

# CONTENTS

## APPENDICES

## TABLES



## FIGURES

## A. Introduction

With a gross national income per capita of $641 and a Human Development Index of 186 out of 187 countries, Niger is one of the least-developed countries in the world (UNDP 2011). Low levels of education have been an important constraint to economic development since the nation gained independence in 1960. According to the United Nations, the adult literacy rate is 28.7 percent, far below the sub-Saharan Africa average of 62.4 percent, and school enrollment and completion rates in Niger are among the lowest in the region. A concerted government effort has produced substantial gains in primary education in the past decade, yielding an increase in gross enrollment from 32.2 percent in 2000 to 66.6 percent in 2010; however, this success is tempered by a persistent gender gap in enrollment and school completion rates (UNESCO 2011). During the same time period, gross enrollment for males increased from about 38 percent to 73.1 percent, while female enrollment rose from 26.0 percent to 59.8 percent. More telling, the completion rate of primary education in 2010 was only 40.7 percent, with a completion rate of 46.0 percent for boys and 35.1 percent for girls.

In an effort to address some of the education-related challenges facing Niger, the Government of Niger (GoN), the United States Agency for International Development (USAID), and the Millennium Challenge Corporation (MCC) developed the Niger Education and Community Strengthening (NECS) project, which is being implemented by Plan International (Plan) along with key partners, Aide et Action, Readsters and Volontaires pour L'Integration Educative (VIE) Kande Ni Bayra.[1] The project's goal is to improve educational opportunities available to children while strengthening links between local communities and state structures, and will include a variety of activities targeted at raising learning outcomes, engaging the community, and encouraging families to enroll and keep their children in school. Throughout all of these activities, NECS will place a special emphasis on girls and literacy. NECS activities will be implemented in 150 villages located in 11 departments and 20 communes across seven regions of Niger.

The NECS project continues and complements previous efforts to improve the educational outcomes of girls in Niger through a program called IMAGINE, implemented by Plan in collaboration with USAID as part of MCC's three-year Threshold Program in Niger (NTP), which began in 2008.[2] The IMAGINE program consisted of the construction of 68 primary schools with high quality infrastructure and implementation of a set of complementary interventions designed to increase girls' enrollment and completion rates. The complementary interventions were intended to include the design and dissemination of training modules for teachers, promotion of extracurricular activities, provision of teacher incentive awards, and implementation of a mobilization campaign in support of girls' education. The NTP, including IMAGINE, was suspended in December 2009 in the midst of implementation due to a constitutional crisis in Niger. At the termination of project activities after 9 months of implementation, 62 of the 68 IMAGINE schools had been constructed; however, the majority of the complementary activities had not been implemented.

Mathematica Policy Research conducted a rigorous evaluation of the IMAGINE program in 2011, one year after school construction was completed (Dumitrescu et al. 2011). Overall, the evaluation found that IMAGINE had a 4.3 percentage point positive impact on primary school enrollment, no impact on attendance, and no impact on math and French test scores. The program impacts were larger for girls than for boys. For girls, the program had an 8.0 percentage point positive impact on enrollment and a 5.4 percentage point impact on attendance, while there were no significant impacts for boys. Given the

[1] VIE was involved early on with the NECS project.

[2] IMAGINE's official name is "IMprove the education of Girls In NigEr."

interruption of the original IMAGINE program, the GoN and MCC agreed to reinstate a portion of the NTP in July 2012. At the same time, USAID, with its own funds and some funds from the NTP, began funding the NECS project to continue and complement the girls' education activities begun under the NTP. The NECS project has thus been designed to include a revised set of complementary quality-based activities.

The NECS program activities, which will be implemented as a package in targeted villages, have been designed to address two strategic objectives. The first objective is to increase access to quality education. Activities will include borehole construction and maintenance, support for de-worming and general hygiene campaigns by mobilizing school governance structures to promote these initiatives, and promotion of gender-equitable classrooms and student leadership activities. Furthermore, NECS will work to engage the community by supporting school management committees and developing a student mentoring program to foster a healthy school environment and motivate parents to keep their children in school. The second objective is to increase student reading achievement by implementing an ambitious Rapid Reading curriculum, which consists of training and supporting teachers in new methods of early grades teaching as well as developing reading materials in local languages. This curriculum will be implemented in 1st and 2nd grades starting in the 2013-2014 school year.[3] The project also aims to promote a culture of reading by establishing community support for reading and developing an adult literacy program.

Figure 1 is a preliminary logic model that shows how the NECS interventions might affect various targeted groups and outcomes of interest. The interventions are listed in the left-hand column, followed by columns showing the groups targeted by each intervention and outcomes potentially improved. The multiple interventions being implemented by NECS target a variety of groups in the community, including children, teachers, parents and other adults, and school management committees. Combined, these interventions are intended to contribute to improving enrollment, attendance, and learning in the short term, but may also improve other outcomes; in the long run they are expected to contribute to improvements in employment and income.

MCC has selected Mathematica to conduct a rigorous evaluation of the NECS program. The proposed approach will enable us to estimate the impacts of the package of NECS interventions with and without IMAGINE infrastructure. We will also be able to evaluate the longer-term impacts of IMAGINE three years after its completion. The remainder of this design report describes our approach to the evaluation. In Section B, we describe the key research questions the evaluation is designed to answer. In Section C, we describe our evaluation design, which builds on the design for the IMAGINE evaluation, and in Section D, we describe the data collection effort for the evaluation. In Section E, we describe our planned cost analyses, while in Section F, we provide a time line for the key evaluation activities. Finally, in Section G, we identify key challenges and risks and our plans to address them. This draft of the design report reflects revisions from the initial draft based on discussions with MCC, MCA-Niger, Plan, and other stakeholders.

[3] The materials required for implementation in Zarma, Hausa, Tamasheq, Fulfulde and Kanuri language schools were simultaneously developed and were initially introduced in schools during the 2013-2014 school year in Grade 1. A revised Grade 1 curriculum was developed by Readsters and implemented at the end of the 2014-2015 school year and during the 2015-2016 school year.

**Figure 1. NECS and IMAGINE Program Intervention Activities and Outcomes**

| Activity | Group Directly Affected | Outcomes | | |
|---|---|---|---|---|
| | | Short Term | Medium Term | Long Term |
| New girl-friendly schools* | Girls | Enrollment, attendance, learning | Academic performance | Employment and income |
| Textbooks* | Students | Access to textbooks, learning | | |
| Early grade rapid reading in local languages | Teachers, students | Teaching techniques in early grade reading in local languages, reading ability, learning | | |
| Reading materials in local languages | Students, adults in community | Access to local language reading materials, reading ability, learning | | |
| Mentoring program | Students | Enrollment, attendance, drop-out, completion, learning | | |
| Promotion of gender-equitable classrooms | Teachers, school management committees | Enrollment and attendance of girls, learning of girls | | |
| Promote leadership training for student government | Students | Student/teacher relations, student autonomy, self-esteem | Attendance, student engagement, and academic performance | |
| Support school management committees (COGES, AMEs, APEs) | School management committees | Community participation in education | Quality of education and support for education | |
| Adult literacy program | Parents and adults in community | Adult literacy, culture of reading | Children's enrollment, attendance, and academic performance | |
| New boreholes** | Students | Safer drinking water | Illness, attendance, retention | General health, employment, and income |
| Facilitating general hygiene and sanitation | | Hand washing | | |
| Supporting de-worming | | De-worming treatments | | |

**Key Assumptions:**

- Schools are sufficiently functional (for example, in terms of infrastructure and management) to support program interventions
- Adequate supply of teachers with the training and motivation to implement the rapid reading curriculum
- Adequate support from MEN inspectors and pedagogic supervisors in monitoring implementation of the rapid reading curriculum
- Sufficient participation and interest in other program interventions by other key target groups (for example, adults in the community and school management committees)
- No major disruptive events in the targeted villages (for example, famine or political unrest)

* IMAGINE intervention activities

** IMAGINE intervention activities that will be completed under NECS

## B. Research Questions

The evaluation of the NECS program will address five key research questions and related sub-questions for the impact evaluation, which we have grouped by question type, and one key research and related sub-questions for the cost analyses:

**Impact on key outcomes**

1. What is the impact of the NECS program in combination with the IMAGINE program on key educational outcomes?

    a. What is the impact on primary education enrollment?

    b. What is the impact on learning as measured by test scores?

    c. What is the impact on attendance rates?

    d. What is the impact on other measures of education quality including completion, dropout, repetition, and transition?

2. What is the impact of the NECS program alone on these key educational outcomes?

**Impact for different subgroups**

3. Are the impacts different for girls than for boys?

4. Are the impacts different for children from households with different asset levels?

**Sustainability of IMAGINE**

5. Have the previously completed IMAGINE investments been sustainable?

    a. What is the current level of functionality and use of the infrastructure constructed under the IMAGINE program?

    b. Did the IMAGINE program have any lasting impacts on key educational outcomes?

**Cost Analyses**

6. Was the NECS program investment justified from a cost perspective?

    a. What was the cost effectiveness of the program?

    b. What was the cost benefit of the program?

    c. What was the economic rate of return (ERR) of the program?

The first two research questions are intended to assess the effects of NECS on key educational outcomes. They follow directly from the hypothesis that by tackling some of the major obstacles to education in the targeted communities, the NECS program will affect both the quantity and quality of education experienced by children in these communities. The evaluation will enable us to evaluate the impacts of NECS both in combination with the improved infrastructure introduced to the IMAGINE schools in the first NTP (question 1), and as a stand-alone program in schools with existing infrastructure (question 2). Evaluating these impacts separately will provide useful evidence for MCC, the Niger Ministry of Education (MEN), Plan, and other stakeholders on the extent to which improved infrastructure—which can be very costly—mediates the impact of quality-based

interventions. We describe the specific educational outcomes we will evaluate and the data we will collect to measure the impacts in further detail in Section D.

The third and fourth research questions are intended to explore differences in impacts by subgroups defined by gender and household asset levels. Because obstacles to education may be more severe for girls than for boys due to cultural and other reasons, girls in Niger typically experience worse educational outcomes than boys. Improving girls' outcomes is therefore a policy priority for the GoN, and some components of the NECS program such as the promotion of gender-equitable classrooms have been specifically designed to address this. Similarly, obstacles to education may be more severe for children from households with greater poverty levels (as proxied by household assets). We will therefore explore differences in program impacts along both of these dimensions.

The fifth research question pertains to the sustainability of the original IMAGINE program and has two parts. The first part involves examining the presence, functionality, and use of IMAGINE-specific infrastructure (such as high-quality classrooms, toilet facilities, and teacher lodging) in IMAGINE villages at the start of the NECS program, and comparing these elements to those available in non-IMAGINE villages. Besides providing valuable long-term evidence on the sustainability of the IMAGINE program itself, this may inform the interpretation of the NECS estimates (for example, if the impacts estimated for research questions 1 and 2 are similar, it could be because the IMAGINE infrastructure has fallen into a state of disrepair and has little interaction with NECS). The second component involves evaluating the long-term impacts of the IMAGINE program on educational outcomes at the start of the NECS program, three years after the end of IMAGINE. This will allow us to assess whether the finding of limited impacts in the initial IMAGINE evaluation remains or has changed after more time has passed for effects to manifest.

We will conduct a detailed cost analysis to determine whether the NECS program was economically justified (question 6). This includes determining the program's effects on a per-dollar basis (cost effectiveness), comparing potential benefits to costs in monetary terms (cost-benefit analysis), and computing a single summary statistic of the economic merits of the program (the Economic Rate of Return, or ERR). These cost analyses are described in Section E.

## C. Evaluation Design

### 1. Overview of Design

We will use a variant of a random assignment design to provide the most rigorous evidence possible to answer the key research questions. The basic random assignment design, which is considered the "gold standard" in impact evaluation, relies on the random assignment mechanism to ensure that those receiving the program (treatment group) are equivalent to those not receiving it (control group); any subsequent difference in outcome changes between the treatment and control groups can then be credibly attributed to the impact of the program. Random assignment provides a more credible comparison group than with other alternatives because all eligible parties have an equal likelihood of being assigned to and receiving treatment. This is more rigorous than other approaches which must rely on observable data to identify potential comparison groups. This approach also is more rigorous, and more powerful than, a pre-post design which compares changes over time for only those that are treated; which cannot account for other trends that may affect outcomes being measured.

The evaluation design for the NECS evaluation is a variant of the basic random assignment design that builds on the random assignment conducted for the IMAGINE evaluation. Specifically, the

NECS evaluation design involves two rounds of clustered random assignment. The first round, which was already conducted at the end of 2008 for the IMAGINE evaluation, involved randomly selecting IMAGINE treatment villages from a pool of potential recipient villages identified by the MEN based on specific criteria (the remaining villages became the IMAGINE control villages).[4] The second round of random assignment, which we conducted in November 2012, involved randomly selecting some of the IMAGINE control villages to receive NECS. This design took into account the planned program implementation, which entails implementing NECS in *all* of the IMAGINE treatment villages, and an additional group of villages selected from the IMAGINE control villages. Because we selected this additional group using random assignment, we can conduct a rigorous evaluation of NECS as described below.

The two rounds of random assignment have resulted in three groups of villages, defined by IMAGINE treatment status and NECS treatment status (Table 1):

- Group A villages will receive NECS-plus-IMAGINE
- Group B villages will receive NECS-only
- Group C villages will serve as the control group (receiving neither IMAGINE nor NECS)

**Table 1. Groups of Villages Under the NECS Evaluation Design**

| | NECS Treatment | NECS Control |
|---|---|---|
| IMAGINE Treatment | A<br>62 villages | |
| IMAGINE Control | B<br>88 villages[a] | C<br>54 villages |

[a] While group B (NECS-only) consists of 88 villages, one village has been dropped from the evaluation for logistical and security reasons (see Section C.2.); the impact evaluation will therefore include only 87 villages in this group.

These three groups of villages are equivalent as a result of the two rounds of random assignment, except for the effects of IMAGINE and/or NECS.[5] The first round involved randomly selecting villages for IMAGINE, so that group A (all IMAGINE treatment villages) is equivalent to the combined groups B and C (all IMAGINE control villages). The second round involved randomly selecting villages from the IMAGINE control villages for NECS, so that groups B and C are equivalent to one another, while retaining the original equivalence to group A.[6]

[4] Specifically, the MEN identified the pool of potential recipient villages in several steps. First, the MEN selected two regions (Tillabéri and Zinder). Then, within each region, they selected two departments and two communes within each department. (The criteria used to select these regions, departments, and communes are unclear). Then, within each commune, 10 villages were identified that met set criteria. These criteria included the number of school-aged girls in the village, access to water within the village, and proximity to a transportation route. The program was later expanded to additional regions: in each region, departments and communes were selected and eligible villages were identified based on the same criteria as before.

[5] However, because eligible villages were purposefully identified by the MEN using certain criteria, they are not necessarily comparable to other villages in Niger.

[6] There is a subtle distinction in the point in time at which these groups can be considered equivalent. Groups B and C are equivalent through 2012, as their equivalence relies only on the second round of random assignment. Groups A and C are only equivalent through 2008 because their equivalence relies on both rounds of random assignment, the first of which took place at the end of 2008. As we discuss in Section C.3, this distinction has implications for the analysis.

Because the three groups are equivalent, comparing outcomes for individuals in these groups will provide credible and rigorous estimates of the impacts of the program. Comparing outcomes for groups A and C will provide an estimate of the impact of NECS and IMAGINE combined (research question 1), while comparing outcomes for groups B and C will provide an estimate of the impact of NECS alone (research question 2). Because Plan will implement the NECS interventions as a package in the same villages, the design will enable us to evaluate only the impact of the combined package of NECS interventions rather than that of individual components. Therefore, outcomes for a representative sample of individuals from these villages will be measured and compared across treatment and control groups.

## 2. NECS Random Assignment

To be compatible with the first round of random assignment for the IMAGINE evaluation, which involved assignment of villages within communes, we conducted the second round of random assignment of villages within communes as well. That is, we randomly selected a number of villages from the IMAGINE control villages in each commune to receive the NECS program (together with all the IMAGINE treatment villages in that commune). The steps we used to complete the NECS random assignment were as follows:

- **Finalize the list of villages for NECS random assignment.** The villages included in the NECS evaluation are the same 204 villages (in 20 communes) that were identified by the MEN as eligible for the original IMAGINE program using the approach described above. We identified the villages that had actually received an IMAGINE school in each commune based on Plan's data, and removed them from consideration for random assignment.[7] These 62 villages, spread across all 20 communes, will automatically receive NECS, and will form group A (NECS-plus-IMAGINE). The remaining 142 villages, again spread across all 20 communes, were included in the random assignment process through which groups B (NECS-only) and C (control) were determined.

- **Allocate the number of NECS-only villages across communes.** The NECS program intends to serve 150 villages, of which 62 automatically received NECS by virtue of having an IMAGINE school. This implied that we had to select a further 88 villages (out of the 142 non-IMAGINE villages available) using the NECS random assignment procedure to meet program targets. Because random assignment was to take place within commune, we had to allocate these 88 NECS-only villages across communes before conducting the draw.

  We had to satisfy several criteria in conducting these allocations. First, we had to ensure that the total number of NECS villages (including NECS-plus-IMAGINE and NECS-only) met the implementation targets of Plan and Aide et Action—78 villages in the 11 Plan communes, and 72 villages in the 9 Aide et Action communes. Second, we wanted to ensure fairness and perceived fairness in the allocations across communes, which was

[7] We identified these villages based on their actual IMAGINE status (whether an IMAGINE school was constructed) rather than their original IMAGINE random assignment status because USAID wanted to ensure that all actual IMAGINE villages received NECS. Ideally we would have preferred to identify these villages based on their IMAGINE random assignment status because random assignment is what guarantees group equivalence. In practice, this difference affected the categorization of 13 of the 204 villages and has some implications for the analysis, which we discuss in Section C.5.

a key request of the MEN. Third, we wanted to protect the design against the possibility of attrition, by ensuring that our proposed allocation that satisfied the first two criteria also included at least 2 of each type of village per commune.[8]

To meet these criteria, we decided to allocate the 88 NECS-only villages across communes using the overall fraction of villages to be randomly selected for each implementing partner. Specifically, there were 74 villages eligible for random assignment in the 11 Plan communes of which 42 (57 percent) were to be selected. We therefore allocated approximately 57 percent of eligible villages in each Plan commune to receive NECS. We conducted a similar allocation for the Aide et Action communes, allocating 68 percent of villages eligible for random assignment to receive NECS in each commune. Finally, we made minor adjustments to the final allocations to ensure the totals were correct after rounding and that the minimum of 2 villages of each type per commune was attained (Table 2).

[8] If we had (for example) only one control village in a commune and for some reason we were unable to collect data in that village, then we would have to exclude the entire commune from any comparisons involving the control group. This is because there would be no control village in that commune, and the design relies on within-commune assignment.

**Table 2. Allocation of Villages to Research Groups by Commune**

| Region | Commune ID | NECS-plus-IMAGINE Villages | NECS-Only Villages | Control Villages | Total Villages | Implementing Partner |
|---|---|---|---|---|---|---|
| Agadez | 1 | 2 | 2 | 6 | 10 | Plan |
| Diffa | 2 | 2 | 5 | 3 | 10 | Aide |
| Dosso | 3 | 2 | 5 | 3 | 10 | Plan |
| | 19 | 2 | 5 | 3 | 10 | Plan |
| Maradi | 4 | 2 | 6 | 2 | 10 | Aide |
| | 5 | 2 | 7 | 3 | 12 | Aide |
| | 6 | 2 | 7 | 3 | 12 | Aide |
| | 7 | 2 | 5 | 3 | 10 | Aide |
| Tahoua | 8 | 2 | 5 | 3 | 10 | Plan |
| | 9 | 2 | 5 | 3 | 10 | Plan |
| | 10 | 2 | 5 | 3 | 10 | Plan |
| | 11 | 2 | 5 | 3 | 10 | Plan |
| Tillaberi | 12 | 6 | 2 | 2 | 10 | Plan |
| | 13 | 5 | 3 | 2 | 10 | Plan |
| | 14 | 6 | 2 | 2 | 10 | Plan |
| | 15 | 5 | 3 | 2 | 10 | Plan |
| Zinder | 20 | 3 | 5 | 2 | 10 | Aide |
| | 18 | 2 | 6 | 2 | 10 | Aide |
| | 16 | 6 | 2 | 2 | 10 | Aide |
| | 17 | 5 | 3 | 2 | 10 | Aide |
| Total | | 62 | 88 | 54 | 204 | |

- **Conduct random assignment.** We conducted the random assignment at a public meeting in Niamey in November 2012, which was attended by all key stakeholders including the MEN and implementing partners. For each commune, we wrote down the names of all villages that were eligible for random assignment on a separate piece of paper and drew these names randomly out of a bag. The first few villages drawn in each commune were assigned to receive NECS, with the exact number depending on the number of NECS villages allocated to that commune (Table 2).

- **Adjustments to the final list.** After random assignment occurred, it was determined that one of the selected NECS-only villages (in commune number 1) had to be dropped from the NECS program for logistical and security reasons. It was replaced by a village from outside the original list of eligible villages (in commune number 12). Neither the original nor the replacement village will be included in the evaluation, although we will collect data in the replacement village for monitoring purposes. The number of NECS-only villages included in the evaluation is therefore 87 rather than 88.

**Random Assignment Conducted in Niamey, November 2012**
**Mr. Barmou Salifou, Secrétaire General du MEN/A/PLN and Mr. Kalilou Tahirou, Secrétaire General Adjoint du MEN/A/PLN**

## 3. Estimating Overall Impacts

Given the use of random assignment, the basic method to estimate impacts consists of comparing the mean outcomes of the various research groups at end line. However, we intend to use regression models to estimate impacts because these have the advantages of providing greater analytical flexibility, accounting for design characteristics such as stratification by commune, and improving statistical precision through the inclusion of control variables.

We will estimate the impact of NECS-plus-IMAGINE by estimating the following ordinary least squares model (OLS) for the sample of group A (NECS-plus-IMAGINE) and group C (control) villages:[9]

$$Y_{ihj,post} = \alpha + \beta IMAGINE\_NECS_j + \delta_k + \lambda X_{ihj} + \varepsilon_{ihj} \qquad (1)$$

Where $Y_{ihj,post}$ is the outcome for child $i$ in household $h$ in village $j$ at end line; $IMAGINE\_NECS_j$ is a binary indicator that is one if $j$ is a group A (NECS-plus-IMAGINE) village and zero if it is a group C (control) village; $\delta_k$ is a vector of binary indicators, one for each commune $k$; $X_{ihj}$ is a vector of control variables that could be correlated with outcomes (the controls could be at the individual, household, or village level); and $\varepsilon_{ihj}$ is a random error term. The parameter of interest in equation (1) is $\beta$, which gives the estimated average impact of NECS-plus-IMAGINE on the outcome of interest (research question 1). Effectively, equation (1) involves a follow-up comparison of groups A and C that assumes equivalence at the time of the original IMAGINE random assignment (in 2008) and captures the effects of any differences between the groups that have arisen since then. More specifically, because group A villages will have already experienced three years of IMAGINE at the

[9] Some of the outcomes of interest, such as enrollment, are binary in nature. However, we still prefer to conduct estimation using a linear probability (OLS) model in these cases, because of ease of interpretation. Nevertheless, we will investigate the sensitivity of our results to using a logit or probit model that accounts for the binary nature of these outcomes.

start of the NECS program, the parameter $\beta$ can be interpreted as the impact of three years of IMAGINE alone, plus two years of IMAGINE combined with the package of NECS interventions.

Our estimates have to account for the fact that outcomes among individuals in the same village—which is the level of random assignment—are likely to be correlated, because they experience many of the same conditions (such as the same teachers and school environment). We will account for the correlation statistically by clustering the regression error terms at the village level to adjust the standard errors. In addition, because the fraction of group A and group C villages varies by commune, we will weight villages by the inverse of their probability of selection. Otherwise, treatment status could be correlated with commune, which could result in biased estimates. These village-level weights will be combined with the sampling weights that we will compute for households within a village.

Similarly, we will estimate the impact of the package of NECS interventions alone by estimating the following OLS model for the sample of NECS-only villages and control villages:

$$Y_{ihj,post} = \alpha + \beta NECS_j + \delta_k + \pi Y_{ihj,pre} + \varepsilon_{ihj} \qquad (2)$$

This model is almost identical to equation (1), with two main differences. First, the treatment variable is now $NECS_j$, a binary indicator that is one if $j$ is a group B (NECS-only) village and zero if it is a group C (control) village. This implies that villages in group A will not be included in this part of the analysis. Second, the model controls for imputed measures of the baseline outcome, $Y_{ihj,pre}$ as one of the control variables. This will be imputed for all children using baseline individual data. Depending upon the explanatory power of time invariant characteristics in imputing such values, this measure will be more or less individualized. Because the baseline level of the outcome variable is generally highly correlated with the end line level in educational studies, this is likely to help reduce variance and hence improve statistical power. If it is not possible to impute values using individual characteristics, a village level average of children in the village at baseline will be used for each outcome. Equation (2) involves a follow-up comparison of groups B and C that assumes equivalence at the time of the NECS random assignment (in 2012) and captures the effects of any differences between the groups that have arisen since then due to the effects of the NECS program. Conceptually, this model is a more flexible version of one that would look at the *change* in outcomes as the outcome of interest, comparing the change in NECS-only to the change in control over the period of the evaluation (a "difference-in-differences" analysis). Once again, we will cluster the standard errors by village and estimate appropriate weights for the analysis. The parameter of interest in equation (2) is again $\beta$, which gives the estimated average impact of the package of NECS interventions only on the outcome of interest (research question 2).[10]

The inclusion of the imputed NECS baseline outcome in equation (2) is the key reason why the impacts of NECS-plus-IMAGINE and NECS alone are estimated separately rather than in a single model. It is not legitimate to control for the baseline outcomes in the comparison of groups A (NECS-plus-IMAGINE) and C (control) in equation (1) under the RCT design because the two groups in that model are equivalent only at the original IMAGINE randomization in 2008. The two groups may not

[10] In our analysis of each outcome for the NECS-only versus Control and NECS-plus-IMAGINE versus Control comparisons, we will assess whether our results are robust to correcting for multiple comparisons. This will ensure that statistically significant impacts for a particular outcome are not simply statistically significant by chance due to having two comparisons for that outcome (this is known as the multiple comparisons problem, see Schochet 2009).

be equivalent by 2012 because the NECS-plus-IMAGINE schools would have already experienced several years of the IMAGINE program. In fact, the true "baseline" outcomes for the model in equation (1) would be outcomes collected prior to 2008, which, unfortunately, were not feasible to collect during the IMAGINE evaluation. The impacts of NECS-plus-IMAGINE and NECS alone must therefore be estimated separately if we want to include the NECS baseline as a control in the RCT analysis and benefit from improved statistical power.[11]

## 4. Estimating Impacts for In-school Children

Because a key component of the NECS program focuses on learning (particularly the early grade reading component for the 1st and 2nd grades) and is school based, one might expect most of the impacts of NECS on learning to be concentrated on in-school children. NECS partners have therefore expressed a strong interest in estimating the impacts on learning for a sample of in-school children, or for the sample of in-school children in early grades. However, these estimates are problematic because of the potential for selection bias. Specifically, they may result in over- or underestimates of the true effect of the program because other aspects of the intervention may induce systematic differences across research groups in the characteristics of children who enroll in or stay in school. For example, if the program encourages children from more disadvantaged backgrounds to enroll, then one might expect their scores to be lower, which would decrease the resulting impact estimates and dampen our understanding of the true effect of the program. Therefore, while we could conduct additional analyses in which we restrict the estimates in equations (1) and (2) to the sample of in-school children, these estimates would have to be interpreted with caution because of the potential for bias due to selection into enrollment.

An alternative approach to obtain unbiased estimates for the sample of in-school children is to inflate the overall estimates from equations (1) and (2) based on the enrollment rate in treatment villages (NECS-only or NECS-plus-IMAGINE). For example, if the enrollment rate in treatment villages is 80 percentage points, we could divide the impact estimates by 0.8, effectively inflating them by 25 percent. This is known as a Bloom adjustment (Bloom, 1984).[12] The key assumption underlying this adjustment is that the impact on learning for out of school children in treatment communities is zero, which may be plausible given the in-school focus of the NECS reading program—the component of the program that directly targets learning.[13] If this assumption holds, these "treatment on the treated" (ToT) estimates can be interpreted as the impact of being enrolled in a NECS school on all children who experienced the in-school NECS program. Crucially, valid ToT estimates still require the village level estimates (known as "intent to treat", or ITT estimates) from equations (1) and (2) before the Bloom adjustment can be conducted.

[11] Including the NECS baseline as a control in equation (1) would imply a difference-in-differences approach that compares the change in outcomes from NECS baseline to NECS end line in group A to the same change for group C. However, this analysis might not have a causal interpretation because it relies on additional assumptions that are not guaranteed by the randomization.

[12] In terms of regression models, this can also be estimated using an instrumental variables (IV) approach (Imbens and Angrist, 1994). In this approach, the learning outcome is regressed on an indicator for enrollment in a treatment school, and village treatment status is used as an "instrument" to adjust for any selection bias.

[13] NECS might still have impacts on the test scores of out-of-school children despite the school-based focus of the reading component. For example, there could be positive spillovers if enrolled siblings share reading materials with non-enrolled siblings, other components of NECS such as adult literacy training could affect out-of-school children in the community, or a child may not be currently in school but could enroll and benefit from the program for some period. These possible impacts are an important caveat to the validity of the adjusted estimates.

## 5. Additional Analyses

In addition to the basic impact estimates described above, we will conduct several additional analyses:

- **Estimating impacts for subgroups.** Exploring the variation in impacts by subgroups is of interest to the evaluation. Key subgroups include those defined by gender (research question 3) and by household asset levels (research question 4). The impacts for a particular subgroup can be evaluated simply by restricting the sample used to estimates equations (1) and (2) accordingly, or by including appropriate interaction terms in these equations. We will also explore variation in impacts by other subgroups of interest, such as those defined by the age or schooling status of the child.

- **Evaluating the sustainability and longer-term impacts of IMAGINE.** We will be able to evaluate the sustainability and longer-term impacts of IMAGINE two years after the end of the program by comparing IMAGINE treatment (group A) to IMAGINE control (groups B and C combined) villages using our baseline data. We will conduct these estimates using a regression model analogous to equation (1), only replacing the indicator $IMAGINE\_NECS_j$ with $IMAGINE_j$, a binary indicator that is one if village $j$ was assigned to the IMAGINE treatment group. For outcomes that are measured at the school level—specifically, school infrastructure—we will estimate similar models at the school level. By comparing school infrastructure and educational outcomes, we will be able to assess whether the infrastructure investments were sustained over time, and whether the IMAGINE program had longer-term effects on key educational outcomes that did not manifest themselves at the time of the IMAGINE evaluation (research question 5).[14]

- **Accounting for differences between IMAGINE assignment status and actual IMAGINE status.** During the IMAGINE project, IMAGINE random assignment was not adhered to in 13 villages in 5 communes. Specifically, 8 villages were assigned to IMAGINE treatment while no IMAGINE school was built, 2 were assigned to IMAGINE control while an IMAGINE school was built, and 3 villages were not included in the IMAGINE random assignment but had an IMAGINE school built.

  These changes affected the villages included in our research groups for the NECS evaluation because we identified the villages eligible for NECS random assignment based on their actual IMAGINE status and not on their original IMAGINE assignment status to ensure compatibility with implementation plans. There may be a concern that this movement of villages across research groups after the IMAGINE random assignment has disrupted the equivalence of the original IMAGINE treatment and control groups on which the comparison between groups A and C (the NECS-plus-IMAGINE estimates) relies. (This is not an issue for the comparison of groups B and C—the NECS-only

[14] The original IMAGINE evaluation included 178 of the 201 villages included in the IMAGINE random assignment process. (The remaining 23 villages were dropped for various reasons, while 3 villages were purposefully selected to receive an IMAGINE school and were not included in the random assignment.) While we will focus our estimates of long-term IMAGINE impacts on the full sample of villages included in the IMAGINE random assignment, we will also explore the sensitivity of our estimates to restricting to the sample of 178 villages in the original IMAGINE evaluation to ensure comparability to those results.

estimates—because the equivalence of these groups relies only on the new round of NECS random assignment.)

To address this concern, we will investigate the sensitivity of our NECS-plus-IMAGINE impact estimates to dropping the villages which violated IMAGINE random assignment.[15] If these estimates are substantively different from those for the full sample, we will prioritize the former because the assumptions underlying the random assignment design are more likely to be satisfied. For the longer-term IMAGINE evaluation, we will use the original IMAGINE random assignment status to obtain intent-to-treat (ITT) estimates, as was done for the original IMAGINE evaluation. These estimates can be interpreted as the impact of being selected into the IMAGINE treatment group.

- **Accounting for bilingual schools.** MCA-Niger has expressed concern that approximately 10 of the villages in the NECS evaluation include bilingual schools, in which early grade instruction in local languages in the norm. Because local language early grade reading is a major component of the NECS program, comparisons using these schools could be different to those using non-bilingual schools. We will address this issue by controlling for bilingual status in our regression models, and exploring the sensitivity of our results to excluding bilingual schools from the analysis.[16]

- **Accounting for potential selection bias due to unobservable outcomes.** Some of the outcomes of interest—dropout, repetition, and transition rates in particular—are observed only for children who are enrolled in school. Because the enrollment decision itself may be affected by the program, an analysis of these outcomes might be subject to selection bias if the program causes different types of children to enroll in the different research groups. For the comparison between the NECS-only and control groups, we can account for this potential selection bias by estimating impacts for a specific sample, namely children who were enrolled at the NECS baseline.[17] Random assignment ensures that these children are equivalent across the two groups on average so that a comparison of their outcomes at end line is valid. (We will be able to verify this using the NECS baseline.) However, for the comparison between the NECS-plus-IMAGINE group and the control group, enrolled children may already differ at the NECS baseline because of the long-term effects of IMAGINE. For these reasons, the results for these outcomes should be interpreted with caution.

[15] These results can be interpreted as impact impacts of NECS-plus-IMAGINE with sample attrition of the villages in groups A and C which violated IMAGINE random assignment. The number of villages that would be dropped in this manner—five in group A and three in group C—is well within acceptable limits for the equivalence between treatment and control groups to be maintained in a random assignment design. For example, it is within the limits defined by the U.S. Department of Education's What Works Clearinghouse (WWC) research standards for random assignment designs.

[16] We will capture bilingual status in our school survey and will confirm the number of bilingual schools and their distribution across research groups using our baseline data.

[17] As we discuss in Section F, the NECS baseline will be conducted during the summer break, before schools open for the 2013-2014 school year. Baseline enrollment will therefore be captured retrospectively, with respect to the 2012-2013 school year.

## 6. Power Calculations

To determine the size of the effects that we will be able to detect given our anticipated sample size, we computed minimum detectable impacts (MDIs)—the smallest impacts that our design will be able to statistically distinguish from zero. The MDIs depend critically on the sample size (both the number of villages and the number of respondents within each village), assumptions on key parameters (such as the intracluster correlation coefficient and the regression R-squared), the power with which we would like to detect effects (typically 80 percent), and the variance of the outcome (which, for binary outcomes, depends crucially on the baseline level of the outcome). Table 3 shows MDIs for several of the key outcomes of interest. To the extent possible, we calculated these MDIs using parameter estimates obtained from the IMAGINE evaluation.

The MDI for the impact of NECS-plus-IMAGINE on the enrollment rate is 7.9 percentage points (13.5 percent of the expected baseline mean), and that for the attendance rate is 8.1 percentage points (16.7 percent of the expected baseline mean). This suggests that we will only be able to detect relatively large impacts on these outcomes. For test scores—which will make use of the full sample, with scores normalized by age group—we will be able to detect an impact of approximately 0.16 standard deviations. This is within the range of test score impacts that one would typically expect for a relatively successful educational intervention.[18]

The MDIs for the impact of NECS alone are lower than the corresponding impacts for NECS-plus-IMAGINE. This is both because the number of villages contributing to the NECS-only estimates is larger, and because these estimates will include an imputed control variable for the baseline level of the outcome, which we expect to increase the amount of variation in the outcome that is explained by control variables (the regression R-squared in Table 3). However, the MDIs will be higher than originally anticipated for two reasons. First, due to insecurity in the region, it will not be possible to collect end line data in the villages located in the Diffa region. This reduces the control group sample by 3 villages, the NECS-plus-IMAGINE sample by 2 villages, and the NECS-only sample by 5 villages. Second, since we are collecting repeated cross-sectional data, and not longitudinal data, we cannot include a control for the baseline level of the outcome for a given child, which we would expect to substantially increase the amount of variation in the outcome that is explained by control variables, but rather imputed values for the baseline outcome. We will, however, include controls for baseline village characteristics. Given these considerations, the MDIs for the impact of NECS alone are 7.3 percentage points (13.3 percent of the mean) for enrollment, and 7.3 percentage points (14.1 percent of the mean) for attendance. For test scores, the MDI is approximately 0.15 standard deviations, again smaller than the corresponding MDI for NECS-plus-IMAGINE. The MDIs for the long-term evaluation of IMAGINE are of a similar magnitude (these MDIs are for the ITT estimates).

As mentioned earlier, we are also interested in separately analyzing impacts for certain subgroups—for example, those defined by gender and household asset levels. Although the individual sample sizes for these subgroup analyses will be lower than for the full sample, we expect to have only slightly lower power for these analyses (Table 3). This is because the correlation of outcomes within-village implies that the number of villages and not the number of individuals is most critical in determining power. For example, for a subgroup comprising one half of the full sample (such as girls),

[18] The test score MDIs for the ToT estimates are obtained from the MDIs in Table 3, inflated by the enrollment rate in treated villages. For example, if the enrollment rate in NECS-plus-IMAGINE villages is 85 percent, the MDI for test scores will be 0.16/0.85, or 0.19 standard deviations.

the MDIs are only about 5-6 percent higher than for the full sample. For a smaller subgroup comprising one fifth of the full sample (such as children aged between 6 and 7 at end line), the MDIs are about 20-22 percent higher than for the full sample.

**Table 3. Minimum Detectable Impacts for the NECS Evaluation Design**

| | Number of Villages (Number of Children) | | Minimum Detectable Impacts (As Percentage of Baseline Mean) | | |
|---|---|---|---|---|---|
| | Treatment | Control | Enrollment (Percentage Points) | Attendance (Percentage Points) | Test Scores (Standard Deviations) |
| **NECS-plus-IMAGINE** | | | | | |
| Research group | A | C | | | |
| Full sample | 60 | 51 | 7.9 | 8.1 | 0.16 |
| | (4,200) | (3,570) | (13.5%) | (16.7%) | |
| Subgroup (50 percent) | 60 | 51 | 8.4 | 8.6 | 0.17 |
| | (2,100) | (1,785) | (14.3%) | (17.7%) | |
| Subgroup (20 percent) | 60 | 51 | 9.8 | 9.9 | 0.20 |
| | (840) | (714) | (16.6%) | (20.5%) | |
| **NECS-only** | | | | | |
| Research group | B | C | | | |
| Full sample | 82 | 51 | 7.3 | 7.3 | 0.15 |
| | (5,740) | (3,570) | (13.3%) | (14.1%) | |
| Subgroup (50 percent) | 82 | 51 | 7.7 | 7.7 | 0.15 |
| | (2,870) | (1,785) | (14.1%) | (14.9%) | |
| Subgroup (20 percent) | 82 | 51 | 8.9 | 8.9 | 8 |
| | (1,148) | (714) | (16.2%) | (17.2%) | |
| **Long-Term IMAGINE** | | | | | |
| Research group | Treatment | Control | | | |
| Full sample | 65 | 136 | 7.0 | 7.4 | 0.13 |
| | (6,500) | (13,600) | (19.5%) | (10.9%) | |
| Subgroup (50 percent) | 65 | 136 | 7.2 | 7.6 | 0.14 |
| | (3,250) | (6,800) | (9.7%) | (11.2%) | |
| Subgroup (20 percent) | 65 | 136 | 7.7 | 8.2 | 0.15 |
| | (1,300) | (2,720) | (10.4%) | (12.1%) | |

Sources: Authors' calculations using data from the IMAGINE and NECS evaluations to estimate key parameters where possible.

Note: MDIs are for a two-tailed test with 80 percent power and a 95 percent level of significance, and were computed using the following formula:

$$MDI = 2.8 * \sqrt{\rho(1-R_v^2) * \left(\frac{1}{N_T} + \frac{1}{N_C}\right) + (1-\rho)(1-R_i^2) * \left(\frac{1}{rnN_T} + \frac{1}{rnN_C}\right)} * \sqrt{\sigma^2}$$

where $\rho$ is the intracluster correlation coefficient (assumed to be 0.1 to 0.15 for test scores and other outcomes based on IMAGINE and NECS baseline data); $R^2_v$ and $R^2_i$ are the regression R-squared values that indicate the amount of variation explained by controls at the village level and individual level respectively (both assumed to be 0.1 for the impact of NECS-plus-IMAGINE, 0.2 for the impact of NECS alone, and 0.1 for the long-term impact of IMAGINE); $N_T$ and $N_C$ are the village sample sizes for the treatment and control groups; $n$ is the child sample size per village (100, assuming 40 households and 1.5-2.5 eligible children per household depending upon the sample of villages based on IMAGINE and NECS data); and $r$ is the survey response rate (assumed to be 100 percent based on the IMAGINE data). The term $\sigma^2$ is the variation in the outcome, which is one for normalized test scores and equal to $p(1-p)$ for a binary outcome with baseline rate $p$ (assumed to be 74 percentage points for enrollment and 68 percentage points for attendance based on NECS data for control villages for the long term IMAGINE analysis, and assumed to be 55-59 percentage points for enrollment and 48-52 percentage points for attendance based on NECS data for control villages for the NECS analyses).

## D. Data Collection

The primary goal of the impact evaluations is to assess the interventions' effectiveness in influencing the outcomes they were designed to address and to test a hypothesis that investing in the whole NECS package will lead to improvements in not only enrollment and attendance, but also learning. In this section, we begin by defining key outcome variables and their indicators to be included in the analysis. We then describe the design of the questionnaires that will be used to collect the impact evaluation data. Finally, we provide a description of the evaluation sample.

### 1. Outcome Definitions

The research questions presented in Section B suggest the following set of outcomes for the NECS and long-term IMAGINE evaluations (the long-term evaluation of IMAGINE will focus on the subset of outcomes denoted *):

**Enrollment (*).** Child-level enrollment rates will be measured for all children in the sample for the current school year. We will also measure enrollment in the two previous school years to inform our analysis of dropout, repetition, and transition (which we discuss below).

**Attendance.** Daily, weekly, and monthly child-level attendance rates will be measured.

**Learning in Local Language.** Child-level learning will be measured using language tests in Hausa, Zarma, Tamasheq, Fulfulde, and Kanuri. Summary scores will be calculated and converted into standard deviations for each test by normalizing by age and language group, and all scores analyzed together as a single local language test score. In addition to the total reading score, we will compute normalized scores and report separate impacts for each of the five reading domains covered by these tests, namely oral language, letter recognition, word reading, oral reading fluency, and reading comprehension.[19] These domain-specific scores will be informative regarding specific reading skills that are affected by the program.

**Learning in French (*)** Child-level learning will also be measured using a French reading test. Although the NECS reading activity is focused on local languages and does not directly target French, part of the theory of change is that this will provide young children the essential building blocks for literacy, ultimately improving their French language ability too. In addition, other NECS activities are focused more generally on supporting literacy and enhance the quality of instruction, which may improve French test scores. As with the local language tests, in addition to the total reading score, we will compute normalized scores and report separate impacts for each of the five reading domains covered by these tests, namely oral language, letter recognition, word reading, oral reading fluency, and reading comprehension. A summary score will be calculated and converted into standard deviations by normalizing by age group.

**Learning in Math (*).** Child-level learning will also be measured using a math test. A summary score will be calculated and converted into standard deviations by normalizing by age group. The comprehensive nature of the interventions suggests that learning may improve across multiple

[19] In our analysis of scores in multiple different reading domains, we will account for the fact that the impacts in some of these domains may be statistically significant by chance due to the large number of domains considered (this is known as the multiple comparisons problem, see Schochet 2009). Specifically, we will assess whether the results are robust to standard methods that adjust statistical significance levels based on the number of outcomes assessed within reading.

subjects, and therefore testing learning in math is useful. However, because the interventions will not directly touch on math skills in the classroom, this outcome is secondary.

**Completion.** A child of primary school graduation age is considered to have completed primary school if they finish and are promoted from 6th grade. Completion is defined only for children of primary school graduation age in a particular survey round.

**Dropout.** An enrolled student is considered to have dropped out of school if they were enrolled in school at baseline but are no longer enrolled at end line. Dropout is therefore only defined for students who were enrolled at baseline.

**Repetition.** An enrolled student is considered to have repeated school if they repeat any grade at any time during the evaluation time frame. Repetition is therefore defined only for students who were enrolled at baseline.

**Transition.** An enrolled student is considered to have transitioned successfully to 3rd grade if they are promoted out of 2nd grade and enroll in 3rd grade during the evaluation time frame. Transition is therefore defined only for students who were enrolled at baseline, and are enrolled at 2nd grade at any point in the evaluation. This measure is of particular interest because the reading intervention activities will focus on 1st and 2nd grades.

Additional characteristics of the children, households, and schools in the sample will be collected to facilitate the subgroup analyses described in the research questions, for boys compared to girls, for households with different asset levels, as well as other subgroups of interest. The research questions related to the long-term evaluation of IMAGINE also focus on infrastructure. Therefore, we will measure school infrastructure existence, status, and maintenance at the NECS baseline. We will also measure these variables at NECS follow-up, together with other school inputs such as presence of materials in the school as well as school-level outcomes related to the NECS activities, such as training, enrollment, and head counts.

## 2. Questionnaire Design

The research questions will be answered using data collected from households in sample villages at both baseline and end line, and data from schools in these villages at end line only. We will collect repeated cross-sectional household data for the NECS evaluation, drawing a new sample of households at end line (the households will also be different from those included in the IMAGINE evaluation).[20] The household questionnaire obtains information about household characteristics, demographics, parents' attitudes towards education, children's educational outcomes (enrollment and attendance), children's desire to enroll in school, and skills in French, local language reading, and math. At baseline, a village questionnaire will also capture information about school infrastructure based on the observations of the interviewers and questions posed to village leaders (because the baseline will be conducted while schools are closed for the summer break, as discussed in Section F, it will not be

[20] There are several ways to estimate the impacts of the NECS program. Although the original design included the collection of longitudinal data, delays in project implementation meant that many of the children in the baseline sample may not have been exposed to the NECS reading curriculum at the time of end line data collection at the end of the 2015/2016 school year. It was determined that a repeated cross-section of households with children aged 6-12 would provide the best sample for estimating the impacts of NECS.

possible to conduct a baseline school survey). The school questionnaire, which will be implemented at follow-up only, gathers information about school characteristics and includes a school roster to collect information on student enrollment and attendance.[21] School officials are asked to report enrollment and attendance information only for those students whose parents have indicated in the household survey that they attend that particular school. Both the household and school questionnaires were developed and conducted as paper questionnaires. Full versions of the questionnaires and sample assessments are included in Appendices B–F.

**Household questionnaire.** The household questionnaire is based largely on the questionnaire from the IMAGINE impact evaluation. This, in turn, was based on a similar questionnaire used for the BRIGHT impact evaluation in Burkina Faso, which drew heavily from other existing questionnaires widely used in developing countries. They include the Demographic and Health Survey (USAID), Multiple Indicator Cluster Survey (UNICEF), and the Living Standards Measurement Study (The World Bank). Mathematica also consulted USAID's EdDataII: Education Data for Decision Making database for country-specific information and sample assessment questions. Relying on existing questionnaires provides two important benefits. First, because they have been widely and successfully used in similar developing countries, including Niger, we can have confidence in their validity and reliability. Second, using questions that are phrased in the same manner as in other countries allows researchers to more easily compare our findings to those from similar surveys, both in Niger and in other countries. Survey questions were adapted or added, where necessary, to provide more detailed information to answer specific research questions related to the NECS impact evaluation.

The household survey includes reading assessments in French and local languages (Hausa, Zarma, Tamasheq, Fulfulde, and Kanuri) that will be administered to all children in the sample. There are many reading skills that have been measured for different purposes, and existing research shows that several emergent reading skills are particularly important to developing reading comprehension: phonemic awareness, alphabetic principle, concepts about print, writing, and oral language (National Reading Panel 2000, Dickenson et al. 2009). Automaticity in letter recognition, word reading or oral reading, is also of particular importance to a child's ability to read and comprehend (National Reading Panel 2000, Dickinson et al. 2009, Abadzi 2006 and Abu-Hamour et al. 2012). If a child cannot read quickly enough, they will not be able to recall what they just read by the time they complete a passage. Of the domains, oral reading fluency is the strongest predictor of reading comprehension (Kim et al. 2010). In turn, the best predictors of oral reading fluency are oral language and letter recognition (Kim and Pallante 2012 and Dickenson et al. 2009). In other words, once a child knows how to de-code (letter and phoneme identification), oral language (specifically, receptive and expressive vocabulary knowledge[22]) is the strongest predictor of being able to read fluently. While most research in literacy acquisition has been conducted with English readers, those cross cultural studies that do exist find similar—though not always identical—patterns across languages and cultures. In multilingual contexts, reading acquisition in a child's mother tongue and in a second language can be similar if the

[21] Baseline data collection was originally planned to occur during the 2012/2013 school year, before roll-out of NECS program activities, and was to include both household and school questionnaires. However, due to delays in the baseline data collection effort, it will occur just prior to the start of the 2013/2014 school year so that key outcomes may be measured prior to program beneficiaries being touched by program activities. Therefore, the school questionnaire will not be administered at baseline.

[22] Receptive vocabulary is the ability to understand spoken words, while expressive vocabulary is the ability to produce the proper word (typically after being shown an image). See Lesaux et al. (2010) for further details.

spelling-to-sound correspondence is consistent in both languages. The degree of consistency can affect how transferable early reading skills are from one language to the other (Dickinson et al. 2004).

There are therefore a large number of outcomes related to reading ability that could be measured for an evaluation; however, as discussed above, there are a handful of skills that have been found to be the most accurate predictors of reading ability and are therefore particularly useful to measure. Because reading skills can be expected to improve as children get older and receive additional reading instruction, the appropriate skills to measure depends on a child's age and schooling level. For example, if a child is young and unable to read fully, oral language is a strong predictor of future reading ability that is appropriate to measure; for older children, measuring reading comprehension may be appropriate.

Therefore, Mathematica has created reading assessments that focus on these predictive skills, in particular on oral language, letter recognition, word reading, oral reading fluency and reading comprehension. Mathematica has used subtasks from existing assessments as examples a basis for measurement of each reading outcome, including the Early Grade Reading Assessment (Gove and Wetterberg 2011; PHARE 2009) and PreLAS 2000 (Duncan et al. 2000), and has modified the question content and protocols to fit the Niger context. The assessments are short enough to limit respondent burden, tightly linked to the NECS reading intervention, and allow for sufficient variation in each outcome measure. Table 4 below shows the different reading outcomes that will be measured as part of the NECS evaluation. Appendix D presents sample assessments which will be finalized after piloting.

**Table 4. Reading Assessments in French and Local Languages**

| Early Reading Domain | Early Reading Subtask (Outcome) | Description of Questions in NECS Survey |
|---|---|---|
| Oral language | Receptive oral vocabulary knowledge | The child is given simple instructions to be followed. |
| | Expressive oral vocabulary knowledge | The child is asked to identify parts of the body and objects in the environment that the administrator points out. |
| | Listening comprehension | A text is read aloud to the child and questions about the text are posed afterwards. |
| Letter recognition | Timed letter identification | The child is given 60 seconds to identify letter names and/or the sounds. |
| Word reading | Timed familiar word reading | The child is given 60 seconds to read simple common words. |
| Oral reading fluency | Read connected text accurately (number of words read correctly) and at a sufficient rate (number of words read correctly in 60 seconds) | The child is given 60 seconds to read words in connected text. |
| Reading comprehension | Respond to questions about the text they have read | The test administrator asks the child reading comprehension questions for the text the child just read. |

The household survey includes math assessments as well. To create these, Mathematica worked with our local data collection partner, CIERPA, and their education expert to create assessments that would be useful in creating benchmarks by which to evaluate the effectiveness of the program. These assessments were created by modifying those that were used in the IMAGINE evaluation. Questions were added to deal with floor and ceiling effects. Existing math surveys, including the Early Grade

Math Assessment (EGMA), and school textbooks from Niger were used to create questions that touch on the important competencies.

In addition to measuring test scores and many of the other outcomes described is Section D.1, the household survey will also gather detailed household information to enable us to conduct subgroup analyses and use household characteristics as control variables in our regression models to improve precision. The household survey consists of the following modules:

- **Household characteristics.** This module includes information about the head of household, such as religion, language, and education. Information about the household is also included. This information consists of GPS coordinates, construction materials used, available water sources, and proxies for wealth, such as cattle, mobile telephone, or radio.
- **Household listing form.** In this module, the respondent provides a complete list of all eligible children between the ages of 5 and 14 (at baseline) and ages 6 to 12 (at end line) residing in the household. Basic information about these children including relationship to the head of household, sex, age, and school enrollment and attendance is also gathered. Questions in this section also cover child labor and parental attitudes towards the education of the child.
- **Education module.** This module will be administered for all children ages 5 to 14 (at baseline) and ages 6 to 12 (at end line) who attended school at any time during the prior school year. Questions address access to textbooks, distance to school, and attendance. Specific information about the school attended, including interventions such as separate latrines, participation in feeding programs, receiving de-worming treatment, and reasons the parents sent the child to school, is also included.
- **Child opinions.** This module consists of questions that are directed towards all eligible children ages 5 to 14 (at baseline) and ages 6 to 12 (at end line) in the household, regardless of school enrollment. Children are asked directly whether or not they have attended school, whether or not they wish to attend school, whether or not they have experienced violence, and measures of gender equity if they have ever enrolled in school.
- **Reading assessments in French.** This module is administered to all children ages 5 to 14 (baseline) or 6 to 12 (end line), regardless of school enrollment, to measure early reading skills in treatment and control villages. Children are posed questions both orally and with preprinted test booklets. Table 4 outlines the reading skills tested in these assessments, describing the each subtask used to measure the different reading outcomes.
- **Reading assessments in local languages.** A local language reading assessment is administered to all children ages 5 to 14 (baseline) and ages 6 to 12 (end line), regardless of school enrollment. One of five local language tests (Hausa, Zarma, Tamasheq, Fulfulde, or Kanuri) will be administered in each village, and the same test will be administered to all sampled children in the village. Children not speaking the specified local language will not be administered the local language test. The local language assessments test the same competencies at the same level of difficulty to each other, and to the tests in French (Table 4), and are not direct translations.
- **Math assessment.** This module is administered to all children ages 5 to 14 (baseline) and ages 6 to 12 (end line), regardless of school enrollment, to obtain a benchmark of mathematics skills in treatment and control villages. Children are asked to count, shown preprinted cards and asked to identify numbers, count items, indicate which number was

the greater of a pair of numbers, and perform simple addition, subtraction, multiplication and division, and are asked to perform simple word problems. A single test is administered to all children, regardless of age and education. Because the test is simple and aims to establish benchmarks, more skilled children are able to answer the easier questions relatively quickly and do not demonstrate signs of boredom during test administration.

**Village Questionnaire**. The village questionnaire allows us to gain basic information about each village and the schools in each village. Questions will be posed to village leaders at each round of data collection about the schools, and about other big programs that occur in the village that may affect schooling or children in the village (for example, a health project could have effects on enrollment, or education programs) for both baseline and end-line. At baseline, interviewers will also visit schools in which sampled children were enrolled and will collect information on infrastructure and its state of repair based on their observations. This information will inform the long-term IMAGINE evaluation. Collecting infrastructure information by external observation is necessary because the baseline will be conducted while schools are closed for the summer break, as discussed in Section F.

**School questionnaire.** The school questionnaire will be implemented at follow-up only. The draft of this questionnaire was based largely on the questionnaire used for the IMAGINE impact evaluation, which in turn was based on the one used for the BRIGHT impact evaluation in Burkina Faso. Mathematica used the World Bank's Standards Measurement Study School Questionnaire as a model. The questionnaire was updated to cover topics specifically related to the NECS program. The final version of the school questionnaire that will be fielded in the end line survey will be modified during the course of the NECS project to incorporate key questions related to school and teacher practices that may arise. At end line, the school survey will be administered just after the household surveys are completed, on the same day when possible. The survey will enable measurement of several of the key outcomes defined above, and will provide a second measure of some of these outcomes—for example, enrollment and attendance—which we can use to validate the results from the household survey. It consists of the following modules:

- **School information panel.** This module includes general information about the school, such as name, region, commune, and position of respondent.
- **School characteristics.** This module gathers information about the school including enrollment numbers by grade, type of school (public or private), textbook availability, whether the school offers food programs, and whether other outside programs that may affect schooling outcomes are active in the school or community.
- **School physical structure.** This module includes questions about the number of classrooms, construction materials, availability of desks and chairs, water supply, existence of and functionality of latrines, existence of a preschool, and teacher housing.
- **School personnel.** This module asks respondents to provide information about the teachers at the school including number and gender of teachers, training levels, and participation in gender sensitivity training.
- **School register.** This module contains information on all of the children identified in the household survey as enrolled at this particular school. The first part of the register will be completed by the interviewer before arriving at the school, while the second part requires the interviewer to verify enrollment and attendance for each child while at the school.

**Questionnaire development.** The household and school surveys were written in both English and French. The local data collection team collaborated with Mathematica to ensure translations were accurate and that idiomatic expressions or language usage particular to Niger had been incorporated. However, French is rarely spoken in rural villages. Faced with the prospect of surveying people of many ethnic groups in their respective local languages, Mathematica decided that the best approach would be to hire local interviewers representing the diverse ethnic backgrounds in Niger who were fluent in both French and local dialects, and train them to translate the survey questions as they conducted the interviews.

Once the questionnaires were developed, they were tested through a pilot data collection effort for which Mathematica randomly selected 10 villages—five IMAGINE treatment and five IMAGINE control—from those villages eligible for the study that are near Niamey. Our aim was to survey households and schools in these villages to identify potential problems with the survey questionnaires and data collection procedures. The pilot test was conducted in November 2012. The pilot included interviewer training, conducting a census and random household selection in each village, identification of schools, administering household and school surveys, and data entry, cleaning, and delivery. Mathematica participated in the training for the pilot, and held several debriefing phone calls with the data collection firm after the pilot. An additional pilot will be conducted to finalize the subtasks in the reading assessments prior to their fielding in fall 2013.

Follow-up discussions with stakeholders allowed for fine tuning of the instruments. Based on the results of the pilot test and follow-up discussions with stakeholders, several changes were made to the questionnaires. We streamlined both questionnaires by removing questions that were redundant or unnecessary to conduct the impact analysis and by focusing the assessments on the questions that worked best and were continually increasing in difficulty. We also improved instructions to data collectors regarding procedures for administering the assessments. In addition, we determined improvements to our field procedures to allow for better matching between children in the household and school surveys.

**Timing.** The household survey described above will be administered at baseline in the early fall of 2013, just before the start of the 2013-2014 school year. End line will take place three years after the children have been exposed to the NECS interventions at the end of the 2015–2016 school year, and will include household and school surveys.[23] This time line is reasonable to see impacts for several reasons. First, there is sufficient time for full implementation of all project activities, even accounting for startup delays, and for one to two years of exposure. Second, for the reading intervention specifically—which may drive changes in reading scores—it seems reasonable to expect impacts after four to six months of exposure, so end line after three years of initial exposure to the revised curriculum, and one year after exposure to the improved revised curriculum, will ensure there is sufficient time for impacts to manifest.

## 3. Sampling Approach

The sample frame for the NECS evaluation is composed of all households located in the 203 villages that are included in the NECS evaluation. These villages include 62 NECS-plus-IMAGINE

[23] As mentioned earlier, due to delays in implementation, not all children were exposed to all final versions of the NECS interventions during the intended school years.

villages, 87 NECS-only villages, and 54 control villages.[24] The baseline survey sample will be composed of 40 households with school-age children (5–14 years old) selected randomly from each village in the sample frame. Households are defined as groups of people living together in a common physical space for a minimum of 6 of the previous 12 months or intend to live together at least 6 months, working together under the authority of a person called the head of household, and taking their meals together or from the same supply of food. To develop a list of eligible households, data collectors first conduct a complete census of all the households in the village and identify those with school-age children. Following the census, 40 eligible households in each village are randomly selected to be surveyed in a public process. Those households that refuse to participate will be noted and replaced so that a sample of 40 households per village is obtained for the baseline. It is important to have 40 households at baseline to ensure a sufficient sample size at end line after attrition for the impact estimation. All schools serving children in the village (up to three maximum) within a 10-kilometer radius of the village will also be surveyed for that village at follow-up (and will have their infrastructure observed and measured by interviewers at baseline). To determine which schools to survey, interviewers will use information collected during the household surveys to identify schools regularly attended by children from each village. At end line, the data collectors will complete a new census, and will draw a new sample of 40 households with children ages 6 to 12 years in each village.

## E. Cost Analyses

In addition to estimating the impact of NECS (alone or in combination with IMAGINE) on key educational outcomes, we will conduct cost analyses to estimate the overall economic merit of the investment. These cost estimates will allow us to compare the economic merit of the program to that of similar educational interventions elsewhere, as well as other social investments. We will conduct these analyses separately for the NECS-plus-IMAGINE investment as well as the NECS-only investment. This will allow us to determine if one of these two program models was a sounder investment from a cost perspective, which may be informative for future policy decisions in Niger and elsewhere. Below we describe the various types of cost analyses that we intend to conduct.

**Cost-effectiveness analysis.** Impact estimates will be used to assess the effects of the NECS program on key educational outcomes in the relevant unit of measurement (for example, the effect on enrollment in terms of percentage points, or the effect on test scores in terms of standard deviations). A cost-effectiveness analysis is needed to assess these effects on a per-dollar basis. We will estimate the cost-effectiveness of the NECS (or NECS-plus-IMAGINE) program in three steps. First, we will obtain impact estimates for the key educational outcomes—including enrollment, attendance, completion, and test scores—from the impact evaluation. Second, we will estimate the costs associated with providing the NECS program (and the IMAGINE program, where relevant) in recipient villages. Third, we will obtain a cost-effectiveness measure for each outcome by dividing the estimated cost by the estimated impact for the outcome.[25] In the case of enrollment, for example, we will divide the costs by the impact on the percentage of enrolled children. To get a broad sense of the magnitude of these cost effectiveness estimates, we will compare them to cost-effectiveness estimates of other education interventions in the literature.

[24] As mentioned earlier, there were originally 204 villages but one was replaced for logistical and security reasons. While we will collect data in this replacement village for monitoring purposes, it will not be included in the evaluation sample.

[25] Because the impact estimates will yield the impact of the NECS program as a whole (including the full package of interventions), all the cost analyses described here will also pertain to the program as a whole.

**Cost-benefit analysis.** Calculations of cost-effectiveness do not take into account potential benefits other than improved educational outcomes. In addition, these measures simply provide estimates of the cost of achieving given educational objectives; they provide no information about whether or not the monetary benefits of these estimates might outweigh the costs. That information is provided through a cost-benefit analysis, where the potential benefits of the NECS program (or NECS-plus-IMAGINE program) are compared in monetary terms to the costs of the program. Estimating the cost-benefit of the program again involves three steps, similar to those used to estimate cost-effectiveness. First, we will estimate the lifetime discounted monetary benefits of the program. This will be done by taking all benefits of the program into account, monetizing these benefits over the lifetime of the beneficiaries and discounting the estimate so it is comparable to the costs. Second, we will estimate the costs associated with providing the NECS (or NECS-plus-IMAGINE) program, using the same cost estimates we used to estimate cost effectiveness. Third, we will obtain the cost-benefit measure by subtracting the estimated costs from the estimated monetary benefits.

The key difference between the cost-effectiveness and the cost-benefit analyses is that the former produces one estimate per outcome that informs how much it costs to improve a specific outcome by some specified amount, whereas the latter yields a single measure that informs whether the overall monetary benefits of the program are larger than the costs. Accounting for all the benefits of the NECS program will require a careful and systematic exploration of the avenues through which the ultimate economic benefit of the program can be realized and an effort to monetize these benefits.

The expected benefits of improved educational outcomes will be realized by improved productivity in the future. In addition to educational outcomes, there are other outcomes that may result in long-term monetary benefits. For example, if child health is improved as a result of the program, lower morbidity, higher productivity, and more years of productive work may be resulting effects. Our exploration of the program impacts will shed light on the set of additional outcomes for which there are measurable effects. However, we will not be observing the children in our sample at the time they will realize their increased economic benefits—which occurs after entering the labor market—and thus we will not have a direct measure of the monetary benefits. We will therefore obtain indirect measures of returns to schooling and other possible outcomes for use in this analysis, using relevant studies from the literature. Mathematica is currently conducting this exercise as part of a cost-benefit analysis for the BRIGHT program in neighboring Burkina Faso, which targets similar outcomes in a similar context; once complete, that analysis will directly inform the analysis for the NECS program.

**Economic rate of return (ERR).** The cost-benefit analysis of the NECS program (alone or in combination with IMAGINE) will be used to calculate ERRs. An ERR estimate represents a summary statistic that reflects the economic merits of a proposed investment. Conceptually, it is the discount rate at which benefits exactly equal costs of a proposed intervention program. The higher the value of the benefits relative to costs, the higher the ERR. We will estimate the ERR for both the NECS-only and the NECS-plus-IMAGINE programs based on our estimates of program costs and benefits. Again, our ERR analysis will likely be strongly informed by that currently being conducted for the BRIGHT program in Burkina Faso, since the contexts are quite similar.

**Cost Data Sources.** A key component in the cost analyses described above is accurate information on costs, which is then combined with information about program impacts. For the NECS-only analysis, the relevant cost information is the difference in costs between NECS-only and control schools; for the NECS-plus-IMAGINE analysis, the relevant cost information is the difference in costs between NECS-plus-IMAGINE and control schools.

There are two main categories of relevant costs in this context—infrastructure costs (school building, well, toilets, and so on) and non-infrastructure costs (early grade reading materials, implementing staff time, teachers' salaries, textbooks, and so on). For the NECS-only analysis, we can assume that all school infrastructure costs and MEN-funded non-infrastructure costs are similar in the NECS-only and control group by virtue of random assignment.[26] The only difference in (non-infrastructure) costs arises from the costs of the NECS program; to estimate this we will require obtaining cost data from the NECS implementation team (this should include materials, transport, staff salaries, and so on).

For the NECS-plus-IMAGINE analysis, we can use the same estimate of NECS program costs. However, we must also account for differences in infrastructure costs between existing and IMAGINE schools, as well as possible differences in lifespan so that these costs can be depreciated accordingly. Ideally, we need comparable estimates of these infrastructure costs for both IMAGINE and existing schools; however, acquiring them is challenging. There are a few possible strategies to collecting these data, and we will work closely with MCA-Niger to identify which is most feasible. The simplest approach is to collect these data from Plan for IMAGINE schools, and from the MEN for non-IMAGINE schools (for the latter, we could request costs of recently-constructed schools and use this as an approximation for the infrastructure costs of non-IMAGINE schools in our sample). Other approaches—such as having an infrastructure expert visit a sample of IMAGINE and non-IMAGINE schools to assess the infrastructure—are much more costly and have not resulted in reliable cost estimates in similar settings in the past (for example, in our evaluation of the BRIGHT project in Burkina Faso). However, if infrastructure cost data are not available from the MEN, we will explore alternative approaches to collecting them.

## F.. Schedule of Activities

We will collect the quantitative data for the impact evaluation through two rounds of data collection.[27] Baseline quantitative data collection will occur in the early fall of 2013, just before the start of the 2013–2014 school year.[28] Given the planned implementation timeline of NECS intervention, it is necessary to collect these baseline data before the start of the school year to minimize the risk that they are contaminated by program activities. This timing will likely provide valid baseline measures of learning (which would probably not have been impacted by the NECS interventions at this point), as well as other educational outcomes such as enrollment with respect to the previous school year (before NECS was introduced). The follow-up quantitative data collection will be conducted in the spring of 2016, towards the end of the 2015-2016 school year, after three school years of exposure to the majority of program activities.[29] Each round of data collection will be followed by a report, as shown in Figure 2.

[26] This might not be the case if the NECS interventions cause a reallocation of resources by the MEN. We do not expect this to be the case, but will explore this possibility to the extent possible through our planned school survey at follow-up.

[27] After several rounds of discussion with MCC and other stakeholders, it was determined that the optimal use of evaluation resources would involve two rounds of data collection for the quantitative impact evaluation—a baseline and a follow-up round.

[28] Although the school year officially begins in October, enrollment is often not complete until November due to rains and harvest activities. We will therefore conclude baseline data collection during October 2013 at the latest.

[29] Although the school year officially ends in May, schools often close early. We will therefore conclude end line data collection by end-April 2016 at the latest.

**Figure 2. Schedule of NECS Evaluation Activities, by Study Year and Quarter**

| | 2013 | | | | 2014 | | | | 2015 | | | | 2016 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| Quantitative Data Collection and Processing | | | B | | | | | | | | | | | E | | |
| Analysis and Reporting | | | | B | B | | | | | | E | E | | | E | E |

B = baseline, E = end line follow-up

- **Baseline data collection and report.** The NECS baseline data will include a household survey only (because schools will not be open at the time) and serve four purposes. First, for the NECS impact evaluation, it will establish equivalence between NECS-only and control groups. Second, it will improve precision of impact estimates for the NECS-only impact estimates by enabling us to control for baseline levels in a regression framework.[30] Third, the NECS baseline data will be used as an end line for the IMAGINE project and allow estimates of long-term IMAGINE impacts. Finally, some of the data collected will be used by implementation partners as baseline information for monitoring and evaluation indicators. We will produce a baseline report containing initial findings; these will include a description of baseline outcome levels, estimates of long-term IMAGINE impacts, and an assessment of baseline equivalence for the NECS evaluation. The report will also include a description of the NECS design and program logic. We will release a cleaned anonymized dataset of the quantitative data along with the baseline report.

- **End line data collection and final evaluation report.** Mathematica will prepare a final evaluation report after the end line data collection. Data collection is expected to occur at the end of the school year for the 2015–2016 school year, after the NECS program has been implemented for three school years. The report will provide final impact estimates of the NECS interventions and final cost analyses; we will again release a cleaned anonymized dataset of the quantitative data along with the final evaluation report.

## G. Limitations and Challenges

While our design offers the best possible opportunity to provide rigorous evidence to inform the key research questions, it has some limitations and may face certain challenges going forward. These potential limitations and challenges, and our plans to address them, include the following:

- **Limited ability to isolate the impact of specific components of NECS.** Plan intends to implement the NECS interventions in the same villages with approximately the same timing. As a result, our impact evaluation design will only be able to identify the impacts of the *package* of NECS interventions; we will not be able to directly disentangle the impacts of specific components.

- **Limited power to detect small impacts.** As we discussed in Section C.6, our design only has sufficient statistical power to detect relatively large impacts. Unfortunately, we have little scope to improve power because the number of villages available in the evaluation—which is the primary determinant of the statistical power of the design—is

[30] As noted earlier, baseline characteristics might not be balanced at baseline and cannot be used as baseline controls for the NECS-plus-IMAGINE impact estimates because data collected at the NECS baseline might have been affected by IMAGINE.

fixed. Nevertheless, we believe the magnitude of the effects we are able to detect is reasonable given the range of effects that are likely to be policy relevant.

- **Limited ability to assess external validity.** The villages included in the evaluation were purposefully identified by the MEN based on certain criteria, and may not be representative of the typical village in Niger. Therefore the external validity of our results is not clear—one might obtain different impacts if the interventions are implemented in other villages in Niger. It may be difficult to rigorously assess external validity because we lack comparable representative data on household and school characteristics from the rest of Niger. However, we will carefully describe the characteristics of the respondents and schools in our sample to provide context for the results. The qualitative component of the evaluation may also shed light on the characteristics of villages and schools that are key facilitators or barriers to program success.

- **Possible adverse events in Niger.** The unstable regional security situation could negatively affect our ability to implement the evaluation or collect data in certain villages or communes. The greater the number of villages or communes affected, the more detrimental this will be to our ability to statistically detect impacts. In addition, several stakeholders have informed us that they were concerned about the effects of the food crisis in rural Niger. If large-scale migration ensues from either problem, this could pose severe challenges to the NECS program as well as to the impact evaluation. Although these events are beyond our control, we will have to closely monitor the situation so that we can adapt as necessary.

## REFERENCES

Abadzi, Helen. 2006. "Efficient Learning for the Poor: Insights from the Frontier of Cognitive Neuroscience." World Bank, Washington, DC.

Abu-Hamour, Bashir, Mu'tah University, Annmarie Urso, and Nancy Mather. 2012. "The Relationships Among Cognitive Correlates And Irregular Word, Non-Word, And Word Reading." International Journal of Special Education, Vol 27, No: 1, 2012.

Bloom, H. S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review,* vol. 8, 1984, pp. 225–246.

Dickinson, David, Roberta Golinkoff, Kathy Hirsh-Pasek, Susan Neuman, and Peg Burchinal. "The Language of Emergent Literacy: A Response to the National Institute for Literacy Report on Early Literacy." January 12, 2009

Dumitrescu, A., D. Levy, C. Orfield, M. Sloan. "Impact Evaluation of Niger's IMAGINE Program." Final report submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research, September 2011.

Duncan, S.E., and E. De Avila. *PreLAS 2000 Examiner's Manual.* Monterey, CA: CTB/McGraw-Hill, 2000.

Gove, Amber, and Anna Wetterberg. "The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy." Research Triangle Park, NC: Research Triangle Institute, 2011.

Imbens, Guido W., and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects." *Econometrica: Journal of the Econometric Society*, vol. 62, no. 2, 1994, pp. 467–475.

Lesaux NK, Crosson AC, Kieffer MJ, Pierce M. 2010. "Uneven Profiles: Language Minority Learners' Word Reading, Vocabulary, and Reading Comprehension Skills." *Journal of Applied Developmental Psychology*, 2010 Dec 1;31(6):475-483.

National Reading Panel. 2000. *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: National Institute of Child Health and Human Development. Available at http://www.nichd.nih.gov/publications/nrp/report.cfm).

PHARE. « Evaluation Initiale Des Compétences Fondamentales En Lecture-Ecriture Basée sur L'utilisation de L'outil EGRA Adapte en Français et en Arabe au Mali. » Mali: Ministry of Education, November 2009.

Schochet, Peter. "An Approach for Addressing the Multiple-Testing Problem in Social Policy Impact Evaluations." *Evaluation Review,* vol. 33, no. 6, 2009, pp. 539–567.