

Are Short-Term Gains in Learning Outcomes Possible?

Evidence from the Malawi Education Sector
Improvement Project

Salman Asim
Ravinder Casley Gera



WORLD BANK GROUP

Education Global Practice

July 2024

Abstract

This paper presents evidence of the impact of a five-year package of interconnected interventions intended to improve learning environments in eight disadvantaged districts in Malawi. The intervention, which was implemented over five years, provided additional finance to schools to support the hiring of additional teachers and construction of learning shelters to improve class sizes in lower primary, along with constructing classrooms and providing results-based finance to reward improvements in staffing. The interventions were targeted to eight districts with longstanding disadvantages in staffing, learning

environments, and learning outcomes, particularly for girls. Employing administrative data and data from a nationally representative independent sample of public primary schools, the analysis finds that these investments closed the gap in learning outcomes between the targeted districts and the rest of Malawi. There is also suggestive evidence that the program reduced learning gaps between girls and boys. The findings suggest that even in a low-income environment with significant constraints, targeted efforts to reduce class sizes can close district-level gaps in learning.

This paper is a product of the Education Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at sasim@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Are Short-Term Gains in Learning Outcomes Possible?

Evidence from the Malawi Education Sector Improvement Project

Salman Asim* Ravinder Casley Gera†

JEL Classification: I21; I28; C99.

Keywords: Education Quality; Primary School; Education Reform; Econometrics.

*Senior Economist, Education Global Practice, World Bank: sasim@worldbank.org.

†Education Specialist, Education Global Practice, World Bank

Introduction

Despite substantial investments in infrastructure and teachers, many low-income countries still struggle with large class sizes in primary school, particularly in lower grades. The typical primary school in Sub-Saharan Africa has more than 30 pupils per class, but average sizes rise to above 50 in countries including the Democratic Republic of Congo, Tanzania and Chad (Bashir et al., 2018). In Malawi, the context for this study, class sizes are particularly severe with a typical school having more than 125 students to a class (ibid.). These poor conditions stem from the failure of national public education systems to keep up with rapid growth in school enrollments, the result of rapid population growth and rising enrollment rates over more than two decades since the introduction of free primary education by the majority of low-income countries in the 1990s (ibid.).

These poor conditions contribute to low rates of learning in lower grades. In Malawi, fewer than 25 percent of students in Grade 2 achieve minimum proficiency levels in reading according to the Early Grade Reading Assessment (EGRA) conducted by the United States Agency for International Development (USAID) (USAID, 2013). These poor levels of learning in early grades contribute to high rates of repetition, further exacerbating large class sizes, a case of ‘early grade bulge’ (Bashir et al., 2018). In the case of Malawi, repetition rates are typically above 25 percent in lower grades (Ministry of Education, 2022). Faced with these poor conditions, many students simply drop out. Fewer than three-quarters of remain in school until Grade 5 in a number of countries in Sub-Saharan Africa, including Uganda, Burkina Faso, and Mozambique (World Bank, b). In Malawi, around one in four students drop out prior to Grade 5 (Ministry of Education, 2022).

Large class sizes may prove detrimental to learning by reducing the level of attention received from teachers by each student, reducing students’ ability to ask and answer questions and seek guidance, and reducing the likelihood that teachers engage in time-intensive teaching tasks such as marking homework. In Malawi, evidence from the Malawi Longitudinal Schools Survey (see section) suggests that, while primary school teachers are as likely as those in neighboring countries to correct mistakes and give positive reinforcement, they are substantially less likely to set and mark homework and to be available to support students after class, activities which increase in time commitment as a result of large class sizes (Asim and Casley Gera, 2024).

In addition to overall poor conditions in lower primary, many low-income countries have wide variations in conditions between schools, with poorer and more remote districts and sub-district areas typically having larger class sizes than wealthier districts and those closer to capital cities, particularly as a result of inefficiencies in the distribution of teachers (Mulkeen, 2010; Bashir et al., 2018). These inequitable conditions are associated with inequities in learning outcomes, with the least well-equipped districts and schools typically having lower learning outcomes. Evidence from cross-sectional data derived from a range of primary level standardized tests suggests that ‘raising the floor’ on test scores by improving the performance of the lowest-performing students and schools is likely to be the most cost-effective way for low-income countries to raise overall learning levels in the short to medium term (Crouch and Rolleston, 2017). In Tanzania, a recent set of reforms targeted to seven disadvan-

tagged regions, providing pre-primary education, teacher training and mentoring for female students¹, led to an increase of more than one-third in Grade 4 learning outcomes between 2014 and 2016, enabling these regions to overtake the rest of the country (Asim et al., 2019).

There is substantial evidence that reducing class sizes can lead to improved test scores, although the evidence is mixed and dependent on context and the size of classes before and after reductions. In high-income countries, Angrist and Lavy (1999), exploiting a class size limit of 40 in Israel to conduct regression discontinuity analysis, estimate that reductions in class size can induce substantial improvements in test scores for older primary school students (Grade 4 and 5); however, later analysis using more recent data suggested no impacts on learning from the class size reduction (Angrist et al., 2017). Similarly, Kreuger (1999), analyzing longitudinal data from the United States, finds reductions in class size associated with improvement in standardized tests of at least four percentile points; however, Hoxby (2000), conducting further analysis on similar data, finds no significant impacts on learning from class size reduction. Analyzing cross-country data from 47 countries with pupil fixed effects, Altinok and Kingdon (2011) conclude that the negative effects of class size are small and observed in a minority of countries.

Turning to low- and middle-income countries, where class sizes tend to be significantly larger, Muralidharan and Sundararaman (2013) combine experimental and panel data and estimate that the addition of contract teachers to schools in India to reduce pupil-teacher ratios (PTRs) led to an improvement in student learning of 0.15-0.16 standard deviations (s.d.) in math and language tests. In addition, Duflo et al. (2015) conduct an experimental evaluation of a similar intervention in Kenya found that the apparent benefits of class size reduction only for students in the classes taught by contract teachers and not for students in other classes with regular Government teachers who experienced a similar reduction in class size, concluding that improved practices and reduced absenteeism by contract teachers compared to Government teachers was the reason for gains in learning rather than class size reduction.

In recent years, new research has directly explored the question of how large class sizes must be to negatively affect learning outcomes. Datta and Kingdon (2021), examining large-scale test data from India with pupil fixed effects, find that negative effects from large class sizes begin at a size of approximately 40 for science subjects and 50 for non-science subjects, and that India could allow its current national PTR to increase to 40² without negatively affecting test scores.

However, there is a lack of rigorous analysis of the effects of class size reduction focused on countries where class sizes are substantially larger. Evidence from Malawi suggests that PTRs above 90 are associated with significantly lower test scores in Grade 4, equivalent to several weeks' lost learning, even when controlling for a wide range of school, teacher and student characteristics (Asim and

¹ The Education Quality Improvement Programme (EQUIP-Tanzania)

² From a current 22.8

Casley Gera, 2024).

Finally, although the evidence primarily assesses the impact of class size on test scores, student repetition and promotion rates – which are generally closely linked to learning – may also respond to reductions in class size. Walter (2018), assessing evidence from a panel of 20 high-, middle- and low-income countries, estimates that reducing PTRs by improving the allocation of existing teachers between schools could improve student promotion rates by between 0.1 and 4.2 percentage points.

In this paper, we present evidence of the impact of a five-year package of interconnected interventions intended to reduce class size in eight disadvantaged districts of Malawi. The intervention, implemented over five years, constructed classrooms; provided additional finance to schools to support hiring of contract teachers and construction of learning shelters; and provided results-based finance to reward improvements in staffing in lower grades. The interventions were targeted to eight districts with longstanding disadvantages in staffing, learning environments, and learning outcomes, particularly for girls.

Employing administrative data and data from a nationally representative independent sample of public primary schools, we estimate the impacts of the intervention on class sizes, repetition rates, and test scores. We exploit the targeting of the interventions to particular districts to conduct difference-in-difference analysis between schools and students in the targeted districts and those in a comparison group of 17 rural districts (all non-urban, mainland districts which did not receive a similar intervention). We find that the interventions succeeded in reducing PTRs and pupil-classroom ratios (PCRs) in lower primary; reduced repetition rates in lower primary by 0.7 points; and closed the gap in learning outcomes between the targeted districts and the rest of Malawi. We also find suggestive evidence that the treatment reduced the gap in learning between girls and boys. The findings suggest that even in a low-income environment with significant constraints, targeted efforts to reduce class sizes can close district-level gaps in learning.

Our study is one of the first to assess the impact of an interconnected set of interventions to reduce class sizes conducted at large scale. Previous experimental evaluations of interventions to reduce class sizes, such as those as described above, have been conducted on a small scale under pilot conditions.³ While randomized evaluations have historically been considered the ‘gold standard’ for causal estimation of the impact of an intervention on selected outcomes (Cartwright, 2011), there is increasing concern that small-scale experiments can have low levels of external validity. Experiments in which allocation to treatment and control is done randomly, but from a purposively selected overall evaluation sample, may meet the technical definition of a randomized control trial without achieving representativeness of the target population (Deaton and Cartwright, 2018). Furthermore, even RCTs in which both sample selection and allocation to treatment are randomized, typically involve

³ For example, Duflo et al. (2015) employs a sample of 70 schools while Muralidharan and Sundararaman (2013) employs a sample of 100 schools.

interventions conducted on a pilot basis with unique implementation arrangements; these may have an extremely low level of external validity in relation to the implementation of similar interventions at scale through conventional public service institutions (Vivalt, 2020; Pritchett, 2019). For example, the impacts of reducing class sizes through provision of contract teachers observed in Kenya in pilot conditions (Duflo et al., 2015) disappeared when the intervention was implemented by Government rather than by a non-governmental organization (Bold et al., 2018). More generally, evaluation of specific individual interventions is inherently of limited external validity to the context of real-life policymaking, where multiple reforms and interventions are typically conducted simultaneously (Pritchett, 2019). By assessing the impact on learning outcomes of a set of interconnected interventions, implemented at scale, our study is able to provide evidence that is likely to reflect the potential impact of similar interventions in other contexts.

Design

Research Questions and General Hypotheses

Our research question is: can a set of interconnected interventions reduce district-level inequities in school staffing and, as a result, reduce student repetition and improve test scores?

Our general hypotheses are:

1. The provision of the MESIP interventions is expected to improve PTRs in lower primary in the eight disadvantage districts compared to in the 17 comparison districts.
2. The improvement in PTR in lower primary is expected to lead to improvements in student repetition rates and learning outcomes.

Context

Malawi has made impressive achievements in access to schooling since the introduction of free primary education in 1994, but its education system has failed to keep pace with increased enrolments, resulting in overcrowded classrooms and understaffed schools. The average class in Grade 1⁴ has about 150 students; and in Grade 2, about 125 students (Bashir et al., 2018).

These large class sizes reflect weaknesses in the management of Malawi's teacher workforce. Malawi has a barely adequate supply of teachers, with a national PTR of 62:1 (Ministry of Education, 2022). However, as recently as 2016 when MESIP began, the national PTR was significantly higher at 69:1 (World Bank, 2015). Although Malawi produces around 4,000 graduates each year from its standard pre-service training for primary school teachers, the Initial Primary Teacher Education (IPTE). However, in recent years, these teachers have typically waited 1-2 years following graduation before being

⁴ Grades are known as Standards in Malawi, but will be referenced as Grades throughout this paper. Grades range from Grade 1-12. Primary Education is from Grade 1-8, and Secondary Education from Grade 9-12.

hired into the national teaching workforce, slowing the reduction in the national PTR.

Moreover, teachers are poorly distributed between districts and between schools. The most poorly-staffed eight districts have an overall PTR of 68 or more, and school PTRs can vary within a single zone (sub-district) by a factor of ten or more (Ministry of Education, 2020). The result is that more than 2,200 of Malawi's approximately 5,800 schools have pupil-qualified teacher ratios (PqTRs) above 90 in lower primary, and more than 800 of these have PqTRs above 190 in lower primary, double the Government of Malawi's target of 60 (Ministry of Education, 2021).

In addition to being poorly distributed between schools, Malawi's teachers are also poorly distributed within schools, between grades, with severe under-allocation to lower grades. Average PqTRs range from 83.2 in Grade 1 to 52.1 in Grade 5 and 21.6 in Grade 8 ((Ministry of Education, 2021). These misallocations exacerbate large class sizes in lower grades.

Mixed methods analysis conducted during MESIP confirms that these large class sizes constrain Malawi's teachers from maintaining certain positive teaching practices, including setting and marking homework and being available for additional support after class (Asim and Casley Gera, 2024).

Severe shortages of classrooms exacerbate large class sizes. The average school pupil-permanent classroom ratio (PpCR) is 98:1, well above the target of 60:1 (MoE, 2022). Moreover, as with teachers, the available infrastructure is inequitably distributed between schools and more than 500 schools have a PCR of more than 160 in lower primary.

These extremely poor conditions lead to high rates of grade repetition and dropout. Most primary schools in Malawi require students to achieve a passing mark in end-of-year examinations to progress to the next grade. In 2021/22, repetition rates nationwide averaged 25 percent (MoE, 2022b). Only 62 percent of students enrolling in Grade 1 survive to Grade 5, and only 39 percent to Grade 8 (ibid.). According to a recent World Bank report, among 12 sampled countries in sub-Saharan Africa, Malawi's students are the most likely to repeat early grades, and the least likely to survive to Grade 8 (Bashir et al., 2018).

Students who remain in primary school experience poor learning outcomes. At Grade 2 level, fewer than 25 percent of students achieve minimum proficiency levels in the Early Grade Reading Assessment (EGRA) conducted by the United States Agency for International Development (USAID). At Grade 6, fewer than 25 percent of students achieve minimum proficiency levels in the Southern African Consortium for Monitoring Educational Quality (SACMEQ) assessment in Mathematics, placing Malawi near the bottom in the region (USAID, 2013).

Disadvantaged districts. In addition to these overall poor conditions and outcomes, Malawi's public school system has historically been burdened by wide disparities between districts. PTRs and PCRs vary significantly between districts, with certain districts demonstrating consistent disadvantage com-

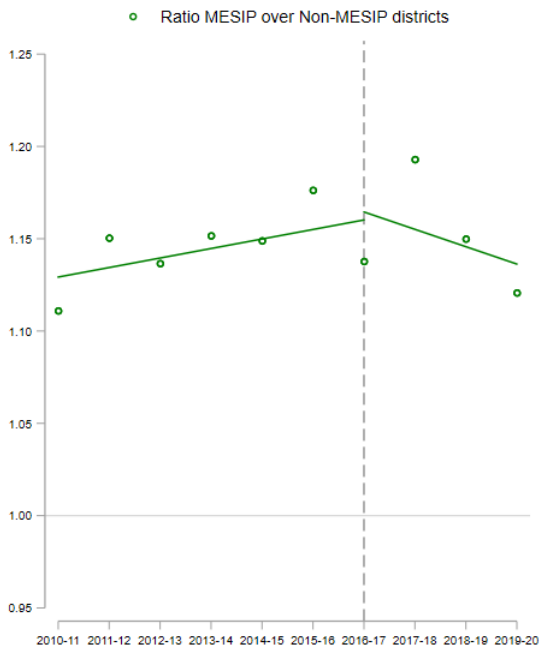
pared to the others. As described in more detail below, the key MESIP interventions were focused in eight districts⁵ selected on the basis of longstanding disadvantage in terms of conditions and learning outcomes.

Figure 1 demonstrates that eight particular PCRs and PTRs in the eight disadvantaged districts were consistently higher in the years leading up to MESIP (2010-2016) than in 17 comparable rural districts (the ratio of PCR/PTR in the eight districts to those in the 17 districts was consistently above 1). See Sample for more details on the construction of the selection of the eight districts and the construction of the comparison group.

⁵ Malawi has 28 districts, the larger of which are subdivided into two or three educational districts for administration of education, for a total of 34 educational districts. For the remainder of this study, we will use the word 'district' to refer to educational districts.

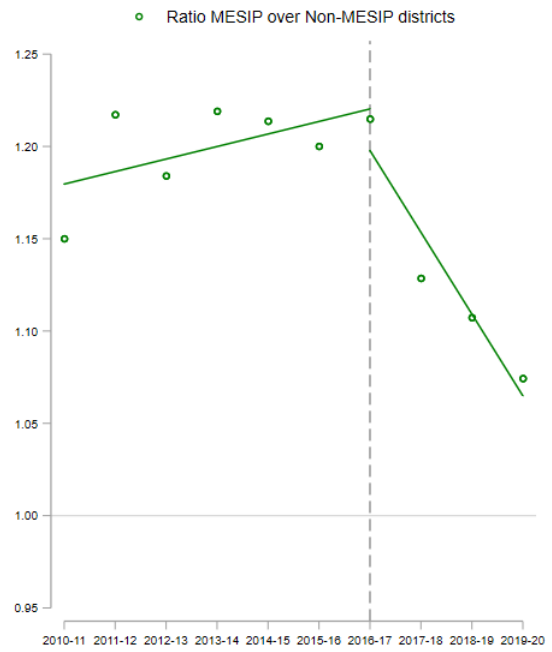
Figure 1: Eight disadvantaged districts: historical trend

Pupil Classroom Ratio (PCR)



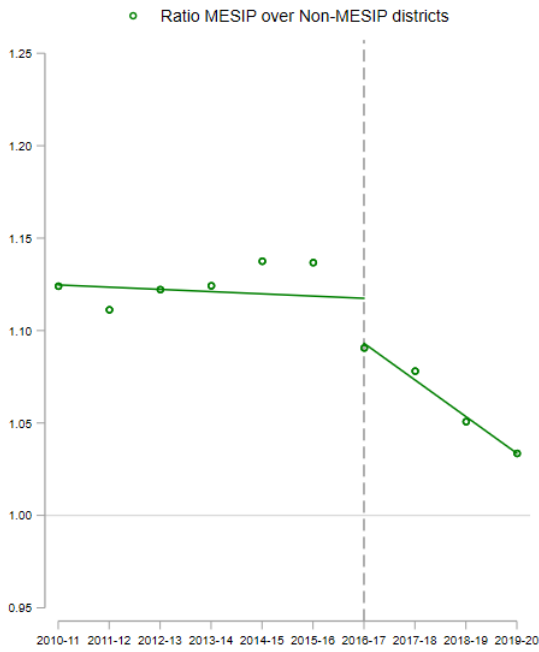
Notes: Figures are WB calculations using EMIS databases for the school years 2010-11 to 2019-20

Pupil Classroom Ratio (PCR) in Lower Primary



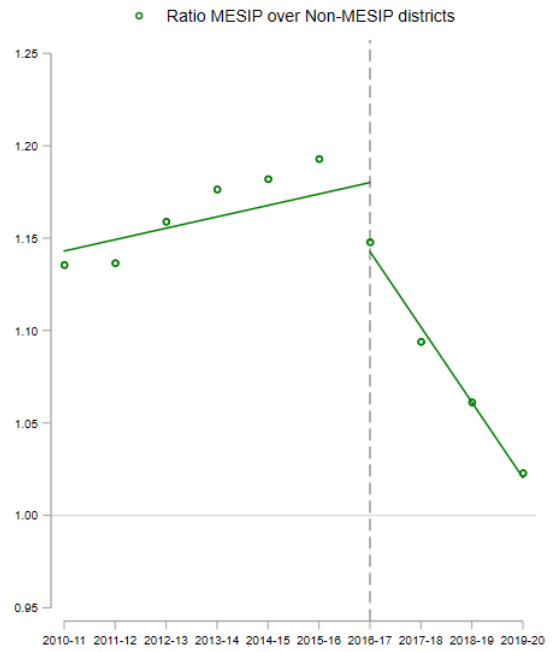
Notes: Figures are WB calculations using EMIS databases for the school years 2010-11 to 2019-20

Pupil Teacher Ratio (PTR)



Notes: Figures are WB calculations using EMIS databases for the school years 2010-11 to 2019-20

Pupil Teacher Ratio (PTR) in Lower Primary



Notes: Figures are WB calculations using EMIS databases for the school years 2010-11 to 2019-20

Intervention

MESIP was a large-scale program of education investment and reforms, implemented by the Ministry of Education of Malawi and financed by the Global Partnership for Education. A sister program, MESIP-Extended, financed by the Royal Norwegian Embassy, extended the MESIP interventions to additional schools. MESIP was implemented from December 2016 to July 2021.

MESIP included a range of interconnected interventions, including:

1. The provision of grants of an average US\$1,761 per year to selected public primary schools, to be used to implement a range of strategies to improve promotion for all students in lower grades and reduce dropout for girls in Grades 6-8;
2. Construction of 500 classrooms and more than 300 sanitation blocks;
3. Results-based financing rewarding Government for improvement in school staffing, specifically (i) reduced PqTRs in Grades 1-2 and (ii) increased female-to-male teacher ratios in Grades 5-8; as well as (iii) reduced repetition rates in Grades 1-4;
4. School Leadership Program providing training to 1200 headteachers, along with their deputies and PEAs, to raise skills in resource management, recordkeeping, teacher management and the creation of inclusive school cultures; and
5. ICT-based community information and engagement, providing schools with either (a) monthly visits from zonal EMIS officers to collect data on key performance indicators, disseminated to communities via SMS and printed report cards; or (b) access to an SMS-based system for community dialogue around education and school conditions, practices and outcomes.

Of these, interventions 1-3 were focused in eight disadvantaged districts of Malawi, with longstanding disadvantages in staffing, learning environments, and learning outcomes, particularly for girls. Interventions 4 and 5 were implemented in a random selection of schools within all districts of Malawi. In this study, we focus on the impacts of the first three MESIP interventions, which were focused in the eight disadvantaged districts and which were expected to lead to reductions in class sizes.

Grants. Malawi already operated a school grants scheme prior to MESIP, called the Primary School Improvement Program (PSIP), which was intended to provide a School Improvement Grant (SIG) of an average US\$950 per year to public primary schools to complete activities recorded in a School Improvement Plan (SIP). However, these SIG were often subject to extreme delays in delivery, often not being received by schools until several months into the school year, and in many cases schools would never receive the full amount (Asim and Casley Gera, 2024). Under MESIP, 800 randomly selected schools within the eight disadvantaged districts received an additional grant, paid on time at the beginning of the school year.⁶ The grants were paid on a per-school basis, with per-learner

⁶ On-time delivery was achieved through payment of funds to schools direct from the MESIP finance, without use of central and local government budgeting and flow of funds processes which are associated with delays and diversion of funds.

top-up payments for schools with more than 1,000 students, and averaged US\$1,761 per year. An additional 400 of these schools were eligible to receive an additional performance-based grant of up to US\$1,200 per year, in exchange for improvements in promotion rates. The first grants were received by schools in May 2017, and following this they were received each year in September-November each year near the beginning of the school year.

The grant finance was accompanied by a set of guidelines specifying particular strategies for the use of the grant finance to improve promotion for all students in lower grades and reduce dropout for girls in Grades 6-8. The strategies included, among others: hiring auxiliary teachers to reduce class sizes (see below); construction of low-cost ‘learning shelters’ to also reduce class sizes where there were insufficient classrooms; provision of rewards for rapidly improving students; zone-level collaboration between teachers to strengthen skills; and provision of materials for remedial classes for low-performing learners. Primary Education Advisers, sub-district education officials, provided support to schools in the implementation of the guidelines.

In addition to the grants and guidelines, a randomly selected subset of 400 schools were selected to receive training on the implementation of the strategies.⁷

In this study, we focus on the two largest expenditure items for schools receiving the grants:

1. *Auxiliary teachers.* A key use of the grant finance by schools was to hire ‘auxiliary’ (contract) teachers to reduce class sizes, particularly in lower primary. Auxiliary teachers (ATs) were qualified teachers who had recently completed pre-service training, but were awaiting deployment into the regular teaching workforce. Under MESIP, they were hired by schools under direct contracts with School Management Committees (SMCs) and allocated to reduce large class sizes in lower primary grades. Although hired by schools directly, ATs were hired on standardized contracts with support to their recruitment and management by PEAs.
2. *Learning shelters.* A second key use of the grant finance was the construction by communities of simple, low-cost classrooms, known as learning shelters. Constructed from brick and with roofs and paved floors, but only partial walls, these shelters used a standardized, low-cost design which was intended to provide adequate shelter for a class at a fraction of the cost of a traditional classroom (around US\$5,000). In addition to shelters, schools also constructed changeroom facilities for the menstrual needs of older female students.⁸

Classroom construction. In addition to the learning shelters produced by schools receiving grants, MESIP supported construction of 500 conventional classrooms, targeted to schools with the

⁷ These schools were randomly selected as part of crossover design for impact evaluation. Two hundred schools receiving only the main MESIP grant, and 200 also eligible for the performance-based grant, received the training. Training lasted one day and was conducted at zone (sub-district) level by district officials.

⁸ Change rooms are simple rooms, often with a sink or other handwashing facility such as a bucket, to provide a dedicated place for girls’ menstrual health.

highest PCR. These classrooms were constructed by large-scale construction firms through centralized contracting. 342 sanitation blocks were also constructed.

Results-based financing. In addition to supporting appointment of auxiliary teachers, MESIP also supported improved staffing and class sizes through results-based finance (RBF). As described above, MESIP provided RBF according to improvement in three areas: early grade PqTRs, upper grade female/male teacher ratios, and lower primary repetition rates. A total of US\$13.5 million of MESIP's total US\$45 million finance was provided through results-based financing, with US\$7 million tied to the targets relating to staffing:

- US\$2 million rewarded the preparation of a strategy and action plan to improve the distribution of teachers between schools.
- US\$2.5 million rewarded the reduction of average PqTRs in Grades 1-2 in the eight disadvantaged districts, from 166:1 to 132:1 (a 20 percent decrease).
- US\$2.5 million rewarded the increase in the average female teacher to male teacher ratio in Grades 5-8 in the eight disadvantaged districts, from 0.31 to 0.34, at ten percent increase.

Theory of Change

We anticipate impacts from the intervention on PTR in schools in the eight MESIP disadvantaged districts, as a result of both the hiring of auxiliary teachers and the improvement in teacher allocation to schools and grades. We expect particular improvement in PTRs in lower primary, as a result of the targeting of the majority of auxiliary teachers to these grades and the priority placed on early grades in the results-based finance.

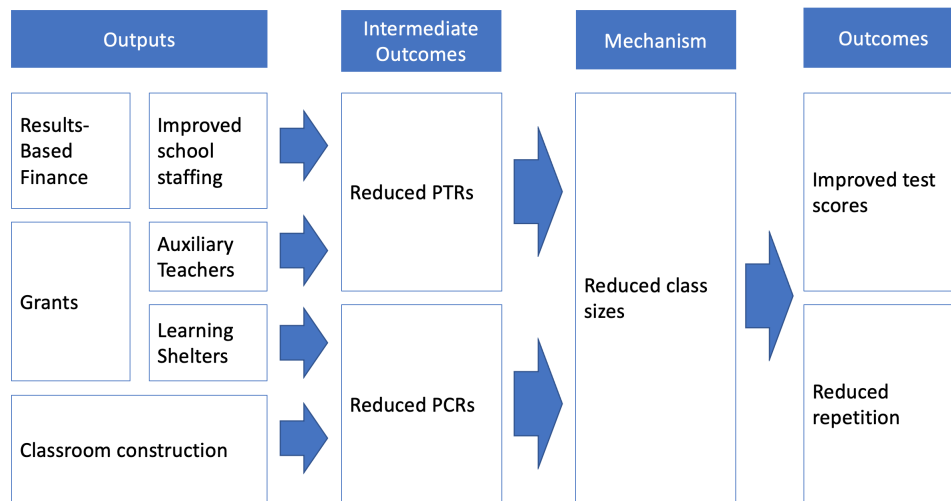
We anticipate impacts from the intervention on PCR as a result of the construction of learning shelters. As learning shelters were assigned for use in lower primary classes, we expect the impacts on PCR to be primarily observed in lower grades.

We anticipate reduced PTRs and PCRs to lead to reduced class sizes.

As a result of these improvements in class sizes, we anticipate improved student learning outcomes and reduced repetition. Repetition is closely associated with learning outcomes in Malawi, with promotion decisions in the eight disadvantaged districts being decided directly through performance in end-of-year school testing. Therefore, we expect improvements in repetition rates and learning outcomes to occur in tandem.

See **Figure 2** for the full theory of change.

Figure 2: Theory of Change



Identification Strategy

We exploit the district targeting of MESIP to use a difference-in-difference approach, comparing the change in outcomes in the eight disadvantaged districts targeted by MESIP with that in the 17 comparison districts. We assume a parallel trend in class sizes and learning outcomes between the disadvantaged and comparison districts, as evidenced by **Figure 1**. We measure the change in outcomes between 2016, the year in which MESIP began implementation, and 2020, the last full year of MESIP implementation.⁹

Outcomes of interest

Our *intermediate* outcomes of interest are school-level PTR and PCR.

PTR: We report PTR at school level using EMIS data. We report PTR rather than PqTR, with no restrictions on the type of teacher.¹⁰ For overall PTR, we sum the total enrollment at each school and divide by the total number of teachers employed at the school. For lower primary PTR, we sum the total enrollment in Grades 1-4 and divide by the total number of teachers employed at the school who teach entirely or partially in these grades.¹¹

⁹ MESIP implementation continued until July 2021; however, most activities were completed by December 2020 and the final months were used to complete activities delayed as a result of COVID-19. Exact measurement date varies according to data source: see Data.

¹⁰ As described, the auxiliary teachers appointed via MESIP SIG were all qualified teachers. Approximately 95 percent of Malawi's primary school teachers are qualified, however, accuracy of school-level administrative data on the qualification status of teachers has not been consistently found to be accurate. Therefore, for simplicity and the reduction of noise, we employ PTR without specifying the qualifications of teachers.

¹¹ The majority of Malawi's primary school teachers are class teachers teaching all subjects to one primary class, however,

PCR: We report PCR at school level using EMIS data. We report PCR rather than PpCR, with no restrictions on the type of classroom.¹² We report PCR school-wide and at lower primary level. For overall PCR, we sum the total enrollment at each school and divide by the total number of classrooms available for use at the school. For lower primary PCR, we sum the total enrollment in Grades 1-4 and divide by the total number of classrooms at the school which are used entirely or partially for these grades.¹³

Our *ultimate* outcomes of interest are student repetition rates and test scores.

Repetition: We report repetition rates at school level, both the overall rate across grades and the lower primary rate across grades 1-4. We calculate repetition rates from EMIS data. The rate is calculated using the number of students in each grade repeating a grade in the most recently completed school year, divided by the total number of students enrolled in that grade for that year; averaged across all grades offered in a school for overall repetition, or across Grades 1-4 for lower primary repetition.

Student test scores: We report student test scores at student level, using data from the MLSS learning assessments, which are norm referenced tests that cover competencies from Grade 1 through Grade 6. We employ data from a longitudinal sample of students who were in Grade 4 at baseline and progressed to upper primary grades during the course of MESIP (see Sample). We report student scores in English, Chichewa and Mathematics. Percentage scores in each subject are adjusted using Item Response Theory and converted to knowledge scores mean-centered at 500 points with a standard deviation of 100 points. For more on the MLSS learning assessments, see Appendix A.

- **Mathematics Knowledge Score:** This is measured by the knowledge score of each student in the mathematics test for each school.
- **English Knowledge Score:** This is measured the knowledge score of each student in the English test for each school.
- **Total Knowledge Score:** This is measured by the knowledge score of each student averaged across both subjects for each school.

in larger schools teachers may teach multiple grades. Our expansive definition of lower primary PTR includes all teachers teaching at least some of their time regularly in grades 1-4.

¹² Although the learning shelters constructed via MESIP SIG are considered permanent classrooms, there is some evidence some were mis-classified as temporary during EMIS data collection; we use PCR to ensure all MESIP-supported and other construction is identified.

¹³ The majority of Malawi's primary school teachers are class teachers teaching all subjects to one primary class, however, in larger schools teachers may teach multiple grades. Our expansive definition of lower primary PTR includes all teachers teaching at least some of their time regularly in Grades 1-4.

Data

We employ both administrative data and data from a large-scale longitudinal survey.

Education Management Information System. The Government of Malawi operates an Education Management Information System (EMIS) with data on all public schools in Malawi. The EMIS data, which is collected primarily through an Annual School Census (ASC) completed by Head Teachers and zonal EMIS officers, includes data on a wide range of indicators including enrollment, dropout and repetition rates, staffing, availability of infrastructure and textbooks. Because it is available for all public schools, the EMIS data is available for the entire sample of public primary schools in all districts of Malawi. For our difference-in-difference analysis, we compare data from EMIS 2016, collected during the 2015/16 school year prior to the start of MESIP; with data from EMIS 2020, collected during the 2019/20 school year towards the end of MESIP implementation.¹⁴ We employ EMIS data for measurement of PTR, PCR, and repetition rates.

Malawi Longitudinal Schools Survey. For test scores, we draw on data from the Malawi Longitudinal School Survey (MLSS). The MLSS is an independent, nationally representative survey that provides data on students, teachers, schools and school communities with the objective of capturing and evaluating the impact of MESIP over 2016-2020. The MLSS is designed to align with the MESIP project cycle to both provide information to the government on implementation and outcomes of the project and to support evaluation of the project. Visits are unannounced.¹⁵ Data is collected through observations and interviews – for example, observations of school and classroom facilities; observation of lessons and teaching practices; interviews with Head Teachers, teachers, and members of community committees; and interviews and testing of students in both a cohort and longitudinal sample (see below). Testing is conducted under classroom conditions in a primarily multiple-choice format supervised by the class teacher as well as MLSS enumerators. For more details of the MLSS instruments and procedures, see Appendix A.

For our difference-in-difference analysis, we compare the data from MLSS baseline and endline. MLSS baseline data collection was conducted in a sample of 559 schools between May and September 2016, just prior to the commencement of MESIP (Asim and Casley Gera, 2024). Endline data collection took place between April 2021 and February 2022, following the completion of the main MESIP activities.

¹⁴ EMIS 2020 data is the last collected prior to COVID-19. COVID-19 introduced significant disruption to schooling (see Asim et al., 2022) and to EMIS data collection; we use the last available pre-COVID data to mitigate these potential confounding factors. This is a conservative approach which is expected to potentially underestimate the full extent of impacts from MESIP activities.

¹⁵ The first visit to each school is unannounced to help capture the real situation of the school in terms of infrastructure, school performance, school and classroom management practices, student and teacher absenteeism and student learning outcomes. If required to complete all instruments, a second visit is made on an announced or pre-scheduled basis.

Sample

District sample

The eight MESIP disadvantaged districts were selected on the basis of being among the worst-off in the nation in terms of PqTR; female/male teacher ratio; shortages of classrooms; and repetition, promotion, and dropout rates in lower primary. The selection was also conducted in such a way as to ensure the inclusion of at least one school from each of Malawi's six sub-regional education divisions. Ultimately, the chosen districts were Lilongwe Rural East, and Dedza from the large Central region; Kasungu from the smaller Northern region; and Chikwawa, Machinga, Mangochi, and Phalombe, all from the largest region, Southern.

For our comparison sample, we first identify the 26 districts not targeted by the main MESIP interventions. We remove the four districts into which these interventions were extended by MESIP-Extended.¹⁶ Although MESIP-Extended supported similar interventions to MESIP, there were differences in sequencing and implementation modality which make it difficult to robustly and simultaneously evaluate both programs.

We also exclude Malawi's four urban districts, as the MESIP interventions were designed for implementation in rural areas. Finally, we also exclude Likoma district, a small island in Lake Malawi with a very particular set of challenges relating to schooling stemming from its particular geography.

With these exclusions, we were left with a group of 17 rural education districts as a comparison group. See **Table 1** for a summary of the selection.

School Sample

For indicators drawn from EMIS, our sample is all public primary schools in the eight disadvantaged districts and the 17 comparison districts (a total of 4,428 schools).¹⁷

For test scores which are drawn from MLSS, our sample is all public primary schools in the eight disadvantaged districts and the 17 comparison districts which are also part of the MLSS sample and for which baseline and endline data is available (a total of 444 schools). The MLSS sample is constructed using stratified probability proportional to size (PPS) sampling was used, with strata defined based on the six educational divisions. From each stratum, a random sample of schools was selected using PPS, using the number of schools in each stratum as measure of size. The PPS process generated a recommended sample size of 700 schools, 571 within the 25 study districts. From these

¹⁶ MESIP-Extended extended the MESIP interventions focused in the eight disadvantaged districts to four more districts which were similarly disadvantaged: Dowa, Mulanje, Nkhhotakota, and Rumphi. MESIP-Extended did not extend the RBF incentives to new districts.

¹⁷ There were 1,856 schools in the eight disadvantaged districts and 2,572 schools in the 17 comparison districts. We include schools which are present in both EMIS 2016 and EMIS 2020 data.

districts, 468 schools were included in the MLSS sample and visited at least once, of which 444 were visited in both the baseline and endline rounds. On average, our sample includes 81 percent of the PPS required schools in each district – an average of 74 percent in the eight disadvantaged districts and 85 percent in the 17 comparison districts (see **Table 2**).

Student Sample

For test scores, we report student level scores from a longitudinal sample of students. As the MLSS is a longitudinal study, it employs both a cohort and longitudinal sample for students. A gender-balanced random sample of 25 Grade 4 students per school (13 girls and 12 boys) is selected at baseline. We then select 15 of these at random (8 girls and 7 boys) for retesting at endline. Students who have dropped out or transferred to other schools are traced to the new schools or their homes and complete learning assessment and a modified version of the student interview. Two students are additionally selected as a reserve sample to replace students who have died, left Malawi, or cannot be traced. We expect to trace 90 percent of the selected students at endline.¹⁸ We compare the baseline and endline scores of each student to calculate student-level difference in learning.

Attrition

There is no school level attrition. At student level, our final sample includes data from 6,940 students of an original longitudinal sample of 7,594 for the sampled schools (91 percent).

Summary statistics

Table 3 shows summary statistics for the key outcomes of interest for the schools in the eight disadvantaged districts and the 17 comparison districts.

Results

Implementation of MESIP

Auxiliary teachers. In total, 478 ATs (362 male and 116 female) were hired by schools in the eight disadvantaged districts (Ministry of Education, 2021). In most cases, these ATs served for 1-2 years before being deployed into the formal teaching workforce. As a custom, ATs were typically (but not exclusively) hired into the same schools in which they had served as ATs before MESIP closed, formalizing the improvements in staffing gained from their hiring.

¹⁸ In addition to this longitudinal sample, at endline the MLSS also selects a cohort sample of 15 students in Grade 4 at endline (8 girls and 7 boys) for testing. In this study, we focus on test scores for the longitudinal sample of students. In addition to testing, students complete a short interview at both baseline and endline.

Learning Shelters. In total, 1,345 learning shelters were constructed by schools in the eight disadvantaged districts. The Government estimates that this construction enabled 79,500 students to move from open-air classrooms to learning in a shelter (Ministry of Education, 2021). The majority of shelters (754) were constructed within the first year of the receipt of grants (World Bank, a). In addition, 542 change rooms were constructed for girls' menstrual needs.

Classroom construction. In addition to learning shelters, MESIP constructed 500 classrooms in the eight disadvantaged districts. The construction was subject to significant delays, with fewer than half (224) constructed by October 2019 and the remainder constructed prior to project closing in July 2021. A total of 342 sanitation blocks were also constructed.

Deployment of regular government teachers. Prior to the implementation of MESIP, there had been severe delays in hiring of teachers into the regular government teaching workforce, with teachers typically waiting two years after graduation from the IPTE before being hired into schools. In 2017, the government deployed 4,900 new teachers. In response to the results-based financing provided under MESIP, this deployment was substantially targeted towards the disadvantaged districts, with 58 percent of these teachers deployed to these eight disadvantaged districts in a bid to reduce the historical understaffing. In 2018, to reduce the shortage of teachers and the waiting period for IPTE graduates, the government conducted a 'double deployment', hiring both IPTE10 and 11 teachers – more than 8,000 in total,¹⁹ compared to the typical 4-5,000. Again, this 'double deployment' was targeted to the eight disadvantaged districts, with around 4,000 teachers allocated to these districts in response to the results-based financing.

Results-Based Financing. The Government fully achieved all three of the staffing-related RBF targets.²⁰ A Primary Teacher Management Strategy was completed and approved by Government for use to improve teacher distribution as well as more generally supporting improved teacher hiring, promotion, transfer and management. The PqTR in Grades 1-2 in the eight disadvantaged districts was reduced from 166:1 to 107:1, a reduction of 36 percent. The female/male teacher ratio in Grades 5-8 in the eight disadvantaged districts was increased from 0.31 to 0.37, an increase of 19 percent (Ministry of Education, 2021).

Difference-in-difference analysis: Intermediate outcomes

Table 4 shows the impact of MESIP on PCR and PTR, drawn from EMIS data. The first row confirms that PCRs and PTRs improved in the comparison group of schools between 2016 and 2020; the second confirms that baseline PCRs and PTRs were higher in the MESIP districts. The third row

¹⁹ Both 2017 and 2018 deployments also included approximately 1,500 Open Distance Learning (ODL) graduates.

²⁰ The Government also fully achieved targets relating to the passage of national strategies and plans to improve female/teacher ratio and reduce repetition rates, but only partially achieved the target relating to repetition (reducing the repetition rate from 23.7 percent to 21.4 percent). In total, the government successfully triggered the release of a total US\$12.2 million out of a maximum US\$13.5 million in results-based financing.

shows difference-in-difference results: PTRs in lower primary decreased significantly more rapidly in MESIP districts, by 14.8 pupils per classroom; overall PTRs decreased by 12.3 pupils per teacher; and PTRs in lower primary experienced a large reduction of 21.4 pupils per teacher.

Difference-in-difference analysis: Outcomes

Repetition

Table 5 shows difference-in-difference analysis of the impact of MESIP on repetition rates. We find that repetition rates increased slightly between 2016 and 2020 in both MESIP and non-MESIP districts. However, we find that the increase in MESIP districts was significantly smaller in MESIP districts in lower primary by 0.7 percentage points.

In order to confirm the relationship between class size reduction and repetition, we explore the dynamics of change in PTR and lower primary repetition rates, exploiting the fact that both PTR and repetition rate are measured using EMIS data which is available on an annual basis. **Table 6** shows the year-by-year change in PTR in both MESIP and comparison districts. A particularly rapid reduction in PTRs in MESIP districts occurs between the 2016/17 and 2017/18 school years, reflecting the commencement of MESIP, the allocation of auxiliary teachers to schools, and the targeting of regular deployment of teachers to the eight disadvantaged districts.

To more clearly assess the relationship between PTR reduction and repetition rates, **Table 7** presents year-on-year, difference-in-difference analyses of repetition rates for a particular subset of treatment and control schools for three time periods: 2017-18 to 2018/19, 2018/19 to 2019/20, and 2019/20 to 2020/21. For treatment schools, for each year-on-year comparison, we limit the analysis to schools in the MESIP districts which had an overall PTR of above 90 in the first year of the comparison, received additional teachers, either auxiliary teachers or new regular government teachers, and as a result had a PTR below 90 in the second year of the comparison. For control schools, we limit to schools in the control districts which obtained one or more additional teachers but did not reduce PTR from above to below 90. This is a lagged analysis, with repetition measured one year following the observed change in PTR. For example, where we report changes in repetition in the period 2017/18-2018/19, the treatment schools are those whose PTR fell below 90 between 2016/17 and 2017/18. This is because repetition rates are established at the end of the academic year when promotion decisions are made, so the first set of promotion decisions reflecting the impact of this change in PTR would be made at the end of the 2017/18 year and reflected in the data for 2018/19.

We find that MESIP reduced repetition rates in lower primary by a large amount - 5.5 percent - between 2017/18 and 2018/19, and that this reduction is statistically significant in comparison to control (difference-in-difference). This suggests that the large reduction in PTR observed in 2017/18 had an impact on promotion decisions made at the end of that year, reflecting improvements in student performance.

Test scores

Table 8 shows the impact of MESIP on test scores. Again, the first row confirms that scores increased in the comparison districts and the second shows that scores in the MESIP districts were below those in the comparison districts at baseline. The third row shows difference-in-difference results. In terms of total scores, students in the MESIP districts achieved an increase in scores 39 points greater than those in the comparison districts, equivalent to several months' additional learning. The average gain in learning for students in the MESIP districts between baseline and endline was 200 points, one-fifth greater than in the comparison districts. The treatment effect is larger than the baseline gap in learning, demonstrating that MESIP enabled students in the eight disadvantaged districts to close the gap with those in the comparison districts. A similar pattern persists with English, Math and Chichewa scores specifically, with Chichewa experiencing the largest impacts from MESIP: students' average Chichewa scores in the eight districts increased by 190 points, one-third more than in the comparison districts.

Gender

A number of aspects of the MESIP interventions were expected to particularly support girls' learning, including the provision of sanitary facilities using grants and the results-based support for improved utilization of female teachers. **Table 9** presents difference-in-difference analysis of the impact of MESIP on test scores, interacted with whether a student is female. Here again, we find suggestive evidence that the effects of MESIP were greater among female students by approximately ten points (in total scores). However, these impacts are not statistically significant as the analysis is not adequately powered to detect significant impacts (there are 2,914 female students in our sample, 52 percent of the total sample.)

Conclusions

Our analysis confirms that, where class sizes are extremely large (above 90 students per class), reductions in class size through hiring of additional teachers and construction of classrooms can lead to reductions in student repetition rates and improvements in test scores. The provision of a package of interventions to reduce class sizes led to significant reductions in lower primary repetition rates and gains of several months' learning in Grade 4 test scores. We also find suggestive evidence that the treatment reduced the gap in learning between girls and boys. These findings have important implications in low-income countries, where class sizes above 90 students remain common in lower primary grades.

Our analysis also confirms that a coordinated package of interconnected interventions, targeted to districts with longstanding disadvantages in conditions and outcomes, can close the gap in learning with other districts. This also has important implications in low-income countries, where inter-district disparities are a key driver of low overall learning outcomes (Crouch and Rolleston, 2017).

In Malawi, with MESIP having substantially addressed inter-district variation in conditions and outcomes, the government has now shifted its investment paradigm to focus on school-level variation. The successor project to MESIP, the Malawi Education Reform Program (MERP), active from 2022 to 2025, scales up the support to hiring of auxiliary teachers and construction of learning shelters introduced under MESIP, providing dedicated finance for auxiliary teachers and construction of low-cost classrooms (using a model adapted from the MESIP learning shelter, with an improved design). Under MERP, this support is targeted specifically to schools with severe overcrowding in lower primary (defined as PqTR or PCR above 90), regardless of district. MERP aims to reduce the share of schools with severe shortages of teachers in lower primary from 70 percent to 30 percent, and the share of schools with severe shortages of classrooms from 61 percent to 30 percent, by 2025.

References

- Altinok, N. and G. Kingdon (2011). New Evidence on Class Size Effects: A Pupil Fixed Effects Approach. Oxford Bulletin of Economics and Statistics 74, 203–234.
- Angrist, J. and V. Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement". The Quarterly Journal of Economics 114, 553–575.
- Angrist, J., V. Lavy, J. Leder-Luis, and S. Adi (2017). "Maimonides' Rule Redux." NBER Working Paper No. 23486.
- Asim, S. and R. Casley Gera (2024). "What Matters for Learning in Malawi? Evidence from the Malawi Longitudinal School Survey".
- Bashir, S., M. Lockheed, E. Ninan, and J.-P. Tan (2018). Facing Forward: Schooling for Learning in Africa.
- Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2018). Experimental evidence on scaling up education reforms in Kenya. Journal of Public Economics 168.
- Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. The Lancet 377.
- Crouch, L. and C. Rolleston (2017). "Raising the Floor on Learning Levels: Equitable Improvement Starts with the Tail". <https://riseprogramme.org/sites/default/files/publications/RISE%20Equity%20Insight%20UPDATE.pdf>. Accessed 1 December 2022.
- Datta, S. and G. Kingdon (2021). "Class Size and Learning: Has India Spent Too Much on Reducing Class Size?" RISE Working Paper 21/059. https://riseprogramme.org/sites/default/files/2021-01/RISE_WP_0059_Datta_Kingdon.pdf. Accessed 1 February 2022.
- Deaton, A. and N. Cartwright (2018). Understanding and misunderstanding randomized controlled trials. Social Science Medicine 210, 2–21.
- Duflo, E., P. Dupas, and M. Kremer (2015). "School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools.". Journal of Public Economics 123.
- Hoxby, C. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation". The Quarterly Journal of Economics 115, 1239–1285.
- Kreuger, A. (1999). "Experimental Estimates of Education Production Functions". The Quarterly Journal of Economics.
- Ministry of Education (2020). Malawi Education Statistics 2019/20.
- Ministry of Education (2021). Project Completion Report: Malawi Education Sector Improvement Project. Mimeo.

- Ministry of Education (2022). Malawi Education Statistics 2021/22.
- MoEST (2015). Education Management Information System 2014-15 [database].
- Mulkeen, A. (2010). Teachers in Anglophone Africa: Issues in Teacher Supply, Training, and Management. World Bank Publications.
- Muralidharan, K. and V. Sundararaman (2013). "Contract Teachers: Experimental Evidence from India." NBER Working Paper No. 19440.
- Pritchett, L. (2019). Randomizing development: Method or madness? Mimeo.
- USAID (2013). "Malawi National Early Grade Reading Assessment Survey: Final Assessment". <https://ierc-publicfiles.s3.amazonaws.com/public/resources/Madagascar%20national%20EGRA.pdf>. Accessed 1 December 2022.
- Vivalt, E. (2020). How Much Can We Generalize from Impact Evaluations? Journal of the European Economic Association.
- Walter, T. (2018). "Misallocation of State Capacity?" PhD Thesis. http://etheses.lse.ac.uk/3852/1/Walter_misallocation-of-state-capacity.pdf. Accessed 1 December 2022.
- World Bank. Malawi Education Sector Improvement Project (MESIP) Mid-Term Review, Aide Memoire. Mimeo.
- World Bank. World Development Indicators: Persistence to Grade 5. <https://data.worldbank.org/indicator/SE.PRM.PRS5.ZS>. Accessed 1 December 2022.
- World Bank (2015). Project Appraisal Document for a Malawi Education Sector Improvement Project.

Tables

Table 1: District Sampling

Division	District	PQTR	Classrooms required	Promotion rate (Std 1-4)	Repetition rate (Std 1-4)	Dropout rate (Std 1-4)	Sample
Central Eastern	Dowa	131.3	908	68.7	25.5	5.8	**
	Kasungu	138.1	1217	69.3	23.1	7.6	MESIP
	Nkhotakota	136.2	746	64.4	29.9	5.7	**
	Ntchisi	107.9	435	64.3	24.7	11.0	Comp
	Salima	127.8	587	64.9	26.8	8.3	Comp
Central Western	Dedza	161.2	1087	66.0	26.7	7.4	MESIP
	Lilongwe City	109.0	713	80.2	18.4	1.4	*
	Lilongwe RE	178.9	817	71.9	17.7	10.5	Comp
	Lilongwe RW	137.7	1016	69.2	22.0	8.8	MESIP
	Mchinji	127.6	969	66.0	28.2	5.7	Comp
	Ntcheu	125.9	760	68.5	27.8	3.7	Comp
Northern	Chitipa	76.4	519	72.8	23.4	3.7	Comp
	Karonga	105.1	799	71.3	26.4	2.3	Comp
	Likoma	49.3	8	69.7	30.3	0.0	*
	Mzimba North	86.8	689	69.5	26.1	4.4	Comp
	Mzimba South	101.3	858	71.0	21.3	7.7	MESIP
	Mzuzu City	87.2	242	80.6	18.6	0.8	*
	Nkhata Bay	83.4	555	70.4	27.3	2.3	Comp
Rumphi	73.7	507	72.4	21.1	6.5	**	
Shire Highlands	Chiradzulu	117.2	516	70.9	28.0	1.1	Comp
	Mulanje	144.4	1137	73.4	23.6	3.0	**
	Phalombe	136.3	499	70.6	22.1	7.2	Comp
	Thyolo	118.4	1258	68.4	27.3	4.3	MESIP
Southern Eastern	Balaka	127.5	650	63.9	28.0	8.1	Comp
	Machinga	173.4	916	68.0	23.8	8.2	MESIP
	Mangochi	191.5	1158	62.7	25.5	11.8	MESIP
	Zomba Rural	145.5	930	69.6	23.8	6.6	Comp
	Zomba Urban	65.0	107	81.1	18.2	0.7	*
Southern Western	Blantyre City	104.2	422	82.2	15.9	1.8	*
	Blantyre Rural	110.4	651	74.7	20.3	5.0	Comp
	Chikwawa	167.4	701	70.7	22.4	6.9	MESIP
	Mwanza	125.7	201	67.8	31.0	1.2	Comp
	Neno	114.5	204	69.6	30.4	0.0	Comp
	Nsanje	141.3	245	70.2	18.4	11.4	Comp

Notes: Yellow highlight denotes MESIP eight disadvantaged districts. Green highlight denotes comparison districts. * denotes urban district/Likoma. ** denotes MESIP-Extended district.

Table 2: School sample

	District	Total schools (EMIS)	PPS sample size	MLSS sample size	Representativeness (%)
MESIP eight disadvantaged districts	Chikwawa	176	21	20	95%
	Dedza	236	39	25	64%
	Kasungu	338	46	33	72%
	Lilongwe Rural West	241	28	28	100%
	Machinga	161	24	11	46%
	Mangochi	259	35	19	54%
	Mzimba South	302	32	27	84%
	Thyolo	179	26	19	73%
	Subtotal	1892	251	182	81%
Comparison 17 districts	Balaka	154	17	15	88%
	Blantyre Rural	157	18	13	72%
	Chiradzulu	88	13	13	100%
	Chitipa	170	26	14	54%
	Karonga	167	19	19	100%
	Lilongwe Rural East	204	28	25	89%
	Mchinji	196	19	11	58%
	Mwanza	45	9	9	100%
	Mzimba North	259	32	23	72%
	Neno	71	9	9	100%
	Nkhata Bay	188	25	20	80%
	Nsanje	104	12	12	100%
	Ntcheu	237	29	26	90%
	Ntchisi	143	18	15	83%
	Phalombe	88	11	10	91%
	Salima	139	16	13	81%
	Zomba Rural	193	19	15	79%
Subtotal	2603	320	262	85%	
Total		4495	571	468	84.9

Table 3: Lower Primary Summary Statistics and T-test differences in MESIP and Non-MESIP districts by year

	MESIP districts			Non-MESIP districts		
	EMIS 2015-16 Mean/(SD)	EMIS 2019-20 Mean/(SD)	Difference T-test/(SE)	EMIS 2015-16 Mean/(SD)	EMIS 2019-20 Mean/(SD)	Difference T-test/(SE)
Enrollment	625.73 (425.60)	647.00 (399.73)	21.27 (13.55)	536.91 (382.41)	542.26 (382.85)	5.35 (10.67)
Teachers	5.68 (4.30)	7.76 (5.07)	2.08 *** (0.15)	5.79 (4.67)	6.61 (4.61)	0.82 *** (0.13)
PCR	159.94 (102.78)	137.12 (87.93)	-22.82 *** (3.14)	131.02 (90.76)	122.99 (84.49)	-8.03 *** (2.45)
PTR	128.46 (73.86)	88.24 (34.15)	-40.22 *** (1.89)	103.97 (60.56)	85.14 (37.79)	-18.83 *** (1.41)
Repetition Rate	0.25 (0.12)	0.26 (0.10)	0.01 *** (0.00)	0.26 (0.11)	0.28 (0.11)	0.02 *** (0.00)
Dropout Rate	0.05 (0.07)	0.05 (0.07)	0.01 ** (0.00)	0.03 (0.05)	0.03 (0.05)	-0.00 ** (0.00)
Schools (N)	2572		-	1856		-

Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table 4: MESIP Impact on PCR and PTR between 2016 and 2020

	(1) PCR	(2) PCR Lower Primary	(3) PTR	(4) PTR Lower Primary
Change from Baseline to Endline (Time)	-4.39** (1.65)	-8.03** (2.45)	-12.07*** (0.86)	-18.83*** (1.41)
Schools in MESIP districts (Treatment)	18.66*** (2.00)	28.92*** (2.98)	14.92*** (1.22)	24.49*** (2.09)
Time x Treatment	-3.34 (2.72)	-14.79*** (3.98)	-12.32*** (1.41)	-21.39*** (2.36)
Constant	105.37*** (1.20)	131.02*** (1.79)	79.65*** (0.71)	103.97*** (1.19)

All schools 4428

Treatment [MESIP districts] 1856

Control [Non-MESIP districts] 2572

Note: * $p < 0.05$, ** $p < 0.01$,

*** $p < 0.001$

Table 5: MESIP Impact on repetition rates between 2016 and 2020

	(1) Repetition Rate	(2) Repetition Rate Lower Primary
Change from Baseline to Endline (Time)	0.016*** (0.003)	0.023*** (0.003)
Schools in MESIP districts (Treatment)	-0.011*** (0.003)	-0.006 (0.004)
Time x Treatment	-0.007 (0.004)	-0.010* (0.005)
Constant	0.251*** (0.002)	0.257*** (0.002)
All schools	4428	
Treatment [MESIP districts]	1856	
Control [Non-MESIP districts]	2572	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Lower Primary Summary Statistics: Schools in MESIP and Non-MESIP districts

	EMIS 2015-16		EMIS 2016-17		EMIS 2017-18		EMIS 2018-19		EMIS 2019-20	
	Non-MESIP Mean/[S.D.]	MESIP Mean/[S.D.]	Non-MESIP Mean/[S.D.]	MESIP Mean/[S.D.]	Non-MESIP Mean/[S.D.]	MESIP Mean/[S.D.]	Non-MESIP Mean/[S.D.]	MESIP Mean/[S.D.]	Non-MESIP Mean/[S.D.]	MESIP Mean/[S.D.]
Enrollment	536.91 [382.41]	625.73 [425.60]	551.66 [393.24]	644.63 [393.87]	549.82 [383.94]	670.43 [410.04]	547.41 [378.71]	665.38 [405.93]	542.26 [382.85]	647.00 [399.73]
Teachers	5.79 [4.67]	5.68 [4.30]	6.08 [4.79]	6.24 [4.48]	6.54 [4.73]	7.34 [4.68]	6.75 [4.82]	7.77 [5.01]	6.61 [4.61]	7.76 [5.07]
PTR	103.97 [60.56]	128.46 [73.86]	98.51 [50.80]	117.62 [60.25]	88.72 [42.24]	97.60 [42.35]	84.47 [35.88]	90.72 [34.75]	85.14 [37.79]	88.24 [34.15]
Schools (N)	2572	1856								

Table 7: Lagged analysis: Repetition Rate in Lower Primary

	Period 2017-18 / 2018-19	Period 2018-19 / 2019-20	Period 2019-20 / 2020-21
Control	0.283 * ** (0.013)	0.283 * ** (0.009)	0.291 * ** (0.010)
Time	0.000 (0.018)	-0.002 (0.014)	-0.057 * ** (0.014)
Treatment	0.006 (0.015)	-0.057 * ** (0.019)	0.023 (0.029)
DiD (Treatment and Time)	-0.055 * * (0.021)	-0.015 (0.024)	-0.014 (0.035)
Control [Non-MESIP]	80	151	164
Treatment [MESIP]	214	48	26

Note: Robust standard errors in parentheses.

Schools in the treatment group are in MESIP districts, received additional teachers in the first year of comparison, and reduced their overall PqTR from above 90 in the first year of comparison to below 90 in the second year.

Schools in control group are in 17 comparison districts, received additional teachers in the first year of comparison, but did not reduce their overall PqTR from above 90 in the first year of comparison to below 90 in the second year.

* p<0.10, ** p<0.05, *** p<0.01

Table 8: MLSS Knowledge scores in 8 MESIP vs 17 Non-MESIP districts

	(1)	(2)	(3)	(4)
	Total	English	Math	Chichewa
Change from Baseline to Endline (Time)	160.58*** (2.45)	176.53*** (2.68)	162.19*** (2.81)	142.25*** (2.86)
Schools in MESIP districts (Treatment)	-27.98*** (2.28)	-26.63*** (2.35)	-30.83*** (2.64)	-27.13*** (3.04)
Time x Treatment	39.26*** (3.82)	29.87*** (4.12)	41.21*** (4.36)	48.10*** (4.52)
Constant	425.20*** (1.46)	418.62*** (1.53)	423.52*** (1.67)	432.82*** (1.92)

Students: In treatment | In control 2255 | 3352

Schools: In treatment | In control 182 | 262

Note: Harmonized scores * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: MLSS Knowledge scores in treatment districts interacted with time and student gender

	(1)	(2)	(3)	(4)
	Total	English	Math	Chichewa
Change from Baseline to Endline (Time)	161.67*** (3.57)	178.65*** (3.89)	164.66*** (4.09)	141.05*** (4.13)
School in MESIP districts (Treatment)	-22.32*** (3.28)	-17.33*** (3.34)	-28.51*** (3.83)	-20.88*** (4.38)
Female student	-5.03 (2.93)	-1.10 (3.06)	-13.23*** (3.34)	-0.46 (3.83)
Time x Treatment	33.81*** (5.58)	21.45*** (5.99)	37.32*** (6.37)	43.03*** (6.61)
Time x Female student	-2.01 (4.92)	-3.99 (5.37)	-4.64 (5.62)	2.40 (5.73)
Treatment x Female student	-10.62* (4.56)	-17.61*** (4.69)	-4.08 (5.27)	-11.85 (6.09)
Time x Treatment x Female student	9.98 (7.65)	15.66 (8.24)	7.05 (8.72)	9.20 (9.05)
Constant	427.79*** (2.09)	419.19*** (2.16)	430.33*** (2.43)	433.06*** (2.71)

Students: In treatment | In control
2255 | 3352

Schools: In treatment | In control
182 | 262

Note: Harmonized scores * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendices

A. Malawi Longitudinal School Survey

The Malawi Longitudinal School Survey (MLSS) collects extensive data on school, classrooms, teachers, Grade 4 students, community members and parents. This Appendix provides summary details. For additional details, see section .

Instruments

The survey contains the following instruments:

1. Observation of school and classroom facilities
2. Lesson observation
3. Head Teacher interview, including details of teachers and committee members; information about their background and procedures; and school information from records
4. Student interview
5. Student learning assessment
6. Teacher interview
7. Teacher knowledge assessment
8. Community interviews with members of the School Management Committee, Parent-Teacher Association Executive Committee, and Mother Group
9. Group Village Headman interview

All instruments were included in all rounds, except the Group Village Headman interview, which was not included in the 2016 phase of baseline.²¹.

The MLSS instruments are based on similar tools used as part of the Service Delivery Indicators (SDI) survey implemented by the World Bank. The SDI instruments were adapted with additional indicators which were appropriate for the Malawian context and/or specific to the MESIP program and related impact evaluations.

Learning assessments: MLSS includes learning assessments in English and Mathematics. These subjects not only allow for capturing of students' literacy and numeracy skills but also allow to test on wide range of cognitive skills. The curriculum of these subjects is relatively standardized across schools. The inclusion of Mathematics and English also makes the test comparable with other standardized international tests. The assessments are targeted to Grade 4 students but contain items aligned with the curricula for Grades 1-6. The test was designed by a psychometrician in collaboration with experts who had prior experience in designing similar tests and were familiar with the primary school syllabus of Malawi. Teachers in government schools were also consulted during the design and formulation of the test items so as to keep the structure of questions as close as possible to

²¹ A Group Village Headman is an intermediary-level official in Malawi's Traditional Authority structure, broadly analogous to a Village Chief

textbooks. Items were developed in reference to international tests and adapted for Malawian context. Student percentage scores are converted to a mean-centred scale, centred at 500, using Item Response Theory.

Sampling

The sampling frame for the MLSS was derived from the most up-to-date list of schools available prior to baseline, the 2015 EMIS (MoEST, 2015).²²

Twenty percent of urban schools that were mainly concentrated in the four major cities of the country (i.e. Blantyre, Lilongwe, Mzuzu and Zomba cities) were randomly selected. For rural schools, a stratified probability proportional to size (PPS) sampling was used, with strata defined based on the six educational divisions. From each stratum, a random sample of schools was selected using PPS, using the number of schools in each stratum as measure of size²³. At the next stage, for districts that had few schools selected using the first round of PPS sampling, random oversampling was conducted to increase the final number of schools in each district to about 24. This oversampling allows district specific analysis. The urban and the rural samples were then combined to form the final survey sample of 924 schools.

As the MLSS is a longitudinal study, it employs both a cohort and longitudinal sample for students. A gender-balanced random sample of 25 Grade 4 students per school (13 girls and 12 boys) is selected at baseline. We then select 15 of these at random (8 girls and 7 boys) for resurvey at endline. Students who have dropped out or transferred to other schools are traced to the new schools or their homes and complete learning assessment and a modified version of the student interview. Two students are additionally selected as a reserve sample to replace students who have died, left Malawi, or cannot be traced. In addition, a new cohort of 15 Grade 4 students is surveyed at endline.

For teachers, a primarily longitudinal sample is used. Ten teachers per school are selected at baseline, using a protocol which ensures representation of lower and upper primary and of female and male teachers while maintaining random selection. All of the sampled teachers are resurveyed at endline if eligible. Teachers who have transferred to new schools are tracked and administered a modified version of the interview, while those who have died, left Malawi, left teaching, or cannot be traced are replaced with teachers who have more recently joined the school or were not selected at baseline.

B. Supplementary Data

Supplementary data associated with this article can be found in the Online Annex at: <https://bit.ly/MESIP-Overall-Evaluation-Annex>.

²² The original sample frame contained 5,738 schools with identifier variables such as division, district and zone. 323 private schools were removed from the sample frame, which leaves the frame with and overall of 5415 primary schools subordinated to government or religious agencies.

²³ Number of schools is used as a measure of stratum size instead of enrollment as the EMIS enrollment data was found unreliable in many instances.