THE REPUBLIC OF UGANDA

Ministry of Education and Sports

Draft report

Measuring Learning Achievement: Baseline for P2-2007

UNITY Project, Uganda

With support from USAID

Report prepared by School-to-School International
for Creative Associates International, Inc.

February 2008

Contents

Tables

Figures

Acknowledgements

The MLA exercise described in this report was conducted by a number of people without whom the quality of the tests and the items, as well as the management of the overall exercise, would not have been possible. In particular, the authors would like to extend their sincerest appreciation to Jeanne Moulton and Sandhya Bandrinath of Creative Associates for overseeing the launching of this exercise and guaranteeing its continued support; Renuka Pillay, Chief of Party of the UNITY Project, for the vigilant attention she paid to the conceptualization and launching of this effort; and to the people listed below, especially the teachers who came in and gave generously of their time to produce what became, in the end, a set of tests of extremely high quality.

A special note of thanks to Martin Opolot who was available at all times, night and day, to expedite training, printing, communications, even emailing late at night in a local cyber café when no other means proved possible.

A heartfelt word of thanks to all.

Dr. Mark Lynd, President
School-to-School International

Names of persons who assisted with the UNITY MLA 2007 exercise

(in alphabetical order)

Name	Title	Institution
Ahabwamukama, Petra	Data Assistant	UNITY Project
Badrinath, Sandhya	Program Officer	CAII
Bale, Remegius	Maths Expert	NCDC
Bananuka, Patrick	Program Officer	UNITY Project
Bertrand, Dr Richard	Independent consultant and President	GAUSS
Dickson, Turyareeba	Research and Policy Analyst	UNITY Project
Drania, Teresa	Teacher	St. Theresa's P/S, Entebbe
Etomet, Joseph	Data Assistant	UNITY Project
Galpern, Andrew	Independent consultant	University of California Berkeley
Herima, Beatrice	Teacher	Nakasero P/S, Kampala
Kabagambe, Richard	Education Planner	Education Planning Department
Kachubya, Frances	Teacher	SOS Hermann Gmeiner, Kakiri, Wakiso
Kasimagwa, Edward	Senior Education Officer	Teacher Education Department
Kizza, Dorothy	Teacher	Buloba C/U, Kampala
Kyagaba, Angela	Language Expert	NCDC
Malinga, Julius	Data assistant	UNITY Project
Moulton, Dr Jeanne	Senior Technical Officer	CAII
Nabagesera, Hamida	Teacher	Namirembe Infants P/S, Kampala
Ndiweru, Mary	Teacher	Bat Valleys, Kampala
Opolot, Martin	Monitoring and Evaluation Officer	UNITY Project
Owere, David	Principal Education Officer	UNEB
Pillay, Renuka	Chief of Party	UNITY Project
Senabulya, Frank	Statistician	Education Planning Department

Acronyms

CAII	Creative Associates International, Inc.
DIF	Differential Item Analysis
GAUSS	Groupe d'analystes universitaires en statistiques sociales
MLA	Measuring Learning Achievement
MOES	Ministry of Education and Sports
PIASCY	Presidential Initiative on AIDS Strategy for Communication to Youth
PMP	Project Monitoring Plan
TDMS	Teacher Development Management System
UNEB	Uganda National Examination Board
UNITY	Uganda Initiative for TDMS and PIASCY
USAID	United States Agency for International Development

Executive summary

The Uganda Initiative for TDMS and PIASCY (UNITY) Project is a USAID-funded initiative. One objective of UNITY is to provide support for the new, thematic curriculum in order to improve teaching and learning at primary school level throughout Uganda. This objective was linked to an indicator in the UNITY Project Monitoring Plan which stated that as a result of UNITY interventions, "at least 70 percent of surveyed children (would) demonstrate higher levels of learning achievement as a result of pre- and in-service training activities."

In order to demonstrate higher levels of learning achievement, the Project, in collaboration with the UNEB and the MOES, initiated a Monitoring Learning Achievement (MLA) exercise through which a sample of pupils would be tested in maths and in language over a 3-year period. The MLA started with a baseline test administered in October 2007, for which tests were developed for P2 pupils in language (in English), maths and performance-based items. The test was administered to 2,325 P2 pupils in 8 districts of Uganda. The results of these tests can be summarized as follows:

1. Tests were of very high level of quality: Numerous statistical procedures indicated very high levels of test and item quality. Test items were also found to be free of gender bias.

2. Test scores revealed significant disparities between regions, pupils' languages, types of schools, status as repeaters, their ages, and their home environments.
    a. *At the regional level*: Pupil scores in both language and maths were the highest in the Western Region, followed by the Central and Northern regions, with the East reporting the lowest scores. Differences between all regions were significant except between the North and the East for the language test, where results were comparable.
    b. *At the district level*: Achievement in Mbarara was significantly higher than all other regions, with the Kumi district reporting the scores significantly lower than all other districts.

c. *By language*: Of the five major languages[1], pupils who reported speaking Runyankole in the home scored highest in both language and maths. Pupils from the Luo and Ateso language groups scored the lowest.
d. *By type of school:* Scores in urban schools were significantly higher than those in peri-urban ones, which were significantly higher than those in rural ones, on all three tests.
e. *By age*: In general, younger pupils performed better than older ones in both language and maths, especially on the language test.
f. *By repeater status*: Non-repeating pupils performed better than repeating ones.
g. *By home environment*: Pupils with books at home and with mothers who read performed better than those without books or whose mothers who do not read.

3. Results between girls and boys were comparable. In language, girls' and boys' scores were comparable in all districts. In maths, their scores were comparable in all districts with the sole exception of Gulu, where boys' maths scores were significantly higher than girls. On the readiness test, girls' performance was the same as boys in all districts on both language and maths items.

4. In most cases, the results on the performance ("readiness") test were similar to those obtained in language and maths. In most cases, scores on the performance-based test (called "readiness test" in this report) were comparable to those on the written language and maths tests. However, some differences were noted. For example, the Mbarara district scored significantly higher than pupils in all other districts on the language and maths tests, though their scores were comparable to those of pupils from Mukono district on the readiness test. And in contrast to the written tests, girls performed the same as boys on maths items of the readiness test across districts, including Gulu.

[1] The five major languages included in this analysis were Ateso, Langi, Luganda, Rukiga and Runyankole.

Based on findings presented in this report, five *methodological recommendations* and four *programmatic recommendations* are proposed. All are summarized here, and are explained in detail in the *Recommendations* section of this report.

The methodological recommendations area as follows:

1. Use tests and items in the forthcoming 2008 and 2009 MLA exercises that are matched to the 2007 tests (not simple translations of the same tests).
2. Ensure careful language selection and translation in future MLA exercises.
3. Revise interview instruments to include key information and categories.
4. Revise selected items, including ones that are too similar and ones that proved too easy for this population.
5. Improve training and monitoring processes.

The report concludes with the following *programmatic recommendations*:

1. Compare results of this assessment to ones conducted by UNEB and others in order to improve the reliability of data collection in future MLA iterations.
2. If characteristics of low-performing pupils remain constant or anomalies are repeated, examine factors associated with these phenomena and explore intervention modalities that might be able to provide assistance to the most vulnerable subgroups.
3. Based on patterns emerging from findings of this and future MLA exercises, identify curricular objectives with which pupils have the greatest difficulties, and provide recommendations concerning the nature of their difficulties and possible courses of remediation.
4. Based on perceptions of Head Teachers and teachers concerning their training in the new curriculum, make possible adjustments to future training content or duration.

Introduction

The Uganda Initiative for TDMS and PIASCY (UNITY) project is a USAID-funded initiative managed by Creative Associates International, Inc. The goal of the UNITY Project (hereafter called "UNITY") is to contribute to quality basic education and the expanded implementation of PIASCY. . One aspect of UNITY focuses on pre-service and in-service teacher training in order to support the Ministry of Education and Sports (MOES) in its implementation of the new national curriculum. UNITY's Project Monitoring Plan (PMP) identifies the measure of the success of these efforts in the following indicator:

> "At least 70 percent of surveyed children demonstrate higher levels of learning achievement as a result of pre- and in-service training activities."[2]

In order to demonstrate that higher levels of learning have indeed occurred as a result of training activities provided by UNITY, Creative Associates International, Inc. engaged the services of School-to-School International (STS) to initiate and coordinate a student testing effort, called Measuring Learning Achievement (MLA), in collaboration with MoES/UNEB, NCDC and EPD.

This report presents the results of the first MLA test administered in October 2007. First, an overview of the MLA design over the life of the project is presented, then methodology used for the 2007 assessment is described – in particular, sampling, test construction, test administration, scoring, data entry and analysis. Findings are then presented, followed by conclusions and recommendations for future MLA exercises. The annexes present a number of tables containing information on item statistics, reliability analysis, sample design, frequencies, and results of inferential analyses.

It should be noted that this report was written with two audiences in mind: decision-makers and educational technicians. The body of the report aims to provide decision-makers with sufficient descriptive information, analyses and recommendations so that they can understand the meaning of results found, on the basis of which sound programmatic improvements can be considered. More

[2] Though the phrase "pre-service and in-service training activities" can refer to training for a variety of populations, the design of this MLA is based on the assumption that this indicator refers primarily to teacher training.

technical details and data tables appear in the report's annexes for use by technicians and decision-makers seeking additional information.

The coordination of the 2007 UNITY MLA was provided by Dr Mark Lynd, President of School-to-School International. International consultant Andrew Galpern assisted with the sampling design and psychometric analysis of drafted test items, and international consultant Dr Richard Bertrand assisted with data analysis and the writing of this report.

MLA design

Over the life of UNITY, it is anticipated that the MLA will consist of three rounds of tests:[3]

In 2007: P2 pupils were tested in language and maths. This test was administered in English and was designed to serve as a baseline for P2 the coming year.

In 2008:

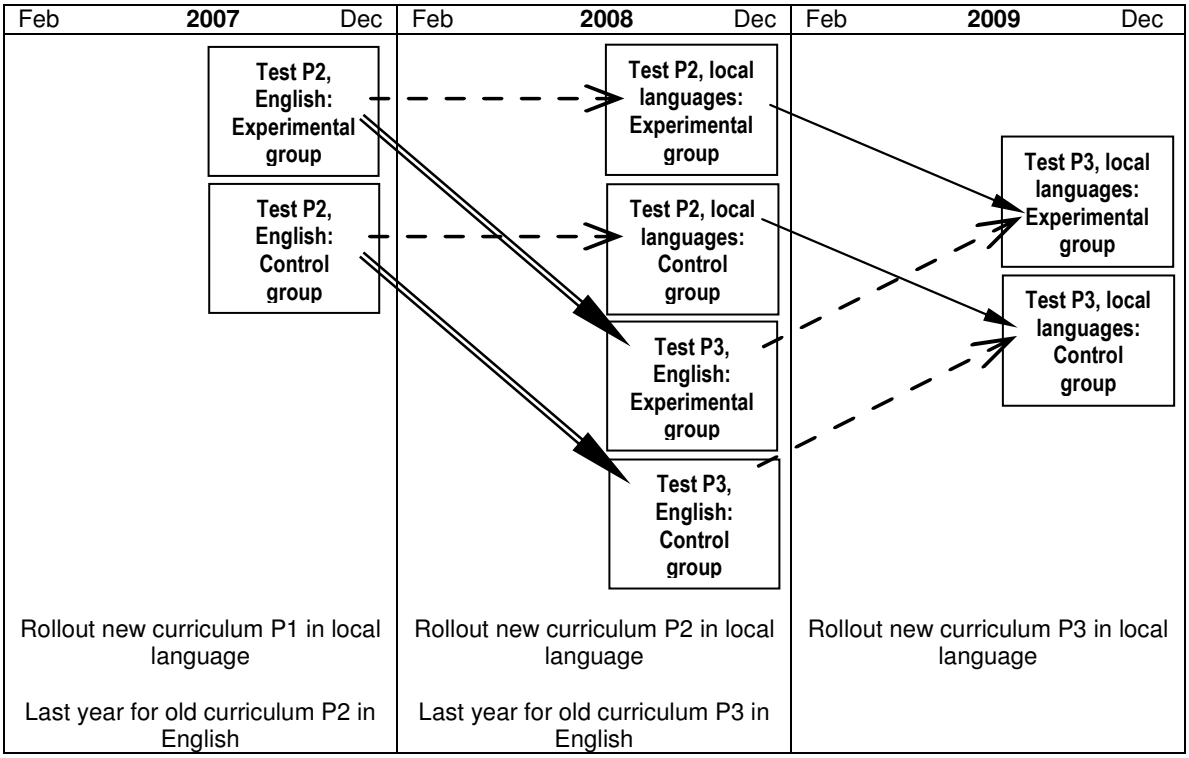
- A new set of P2 pupils will be tested with tests comparable to the ones used in 2007, but this time they will be translated into 4 local languages – Luganda, Luo, Ateso and Runyankole-Rukiga.
- As was done with P2 pupils in 2007, P3 pupils will also be tested in language and maths using a test in English. Again, this will serve as a baseline for P3 in 2009.

In 2009: P3 pupils will be tested in language and maths, again with the P3 test comparable to the one previously administered in English, this time translated into the same 4 local languages as was done for the P2 test.

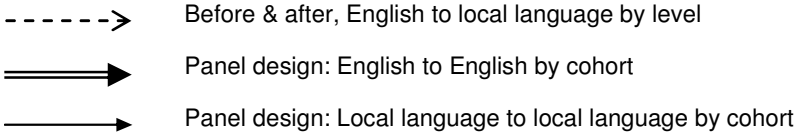
The proposed design can be illustrated as follows in Figure 1:

[3] UNITY and UNEB have discussed and provisionally approved this design; as of this writing, finalization is under discussion.

Figure 1: Proposed MLA design, 2007 – 2009



Legend



The above design has been recommended for the following reasons:

- Two information sources: It provides information in two areas of the curriculum – P2 and P3 – thereby giving decision-makers important information on pupil performance at two different grade levels.
- Panel design: It follows two cohorts – one from 2007 to 2008 and one from 2008 to 2009 – thereby providing insights at the individual pupil level – a rich source of data.[4]
- Margin of security: In the unlikely event that significant problems are experienced in the administration of one test, an additional test will provide sufficient information to ensure that a response to the PMP indicator will be produced.

[4] This cohort model is also called a panel design.

Methodology

Sample

The selection of schools and pupils within schools for this assessment was based on a *stratified 2-stage cluster sampling design*. It is a 2-stage cluster design because sampling was accomplished in two steps: the identification of schools in groups (or clusters), then pupils in those schools were selected. Stratification was based on geographical characteristics, namely region and district. The Eastern, Northern, Western and South Central regions were selected in order to represent the largest populations and language groups, as well as to represent the geographic diversity of the country. Within each region, one urban and one rural district were selected through purposive sampling in order to represent those two conditions; the most remote districts were excluded due to time constraints. Then, within each district, the selection of schools was made according to criteria laid out in the following sections. Once these criteria were established, government schools were selected randomly within the categories specified below. Control (private) schools were selected through convenience sampling. Finally, once in the schools, test administrators selected pupils randomly.

When implementing this sampling design, three criteria were considered key: language, school attributes and sample size. Each of these is examined below:

Language

The main criterion in the selection of schools and pupils to be tested was that of language: since the new curriculum calls for the use of local language as the medium of instruction from P1 to P3, it was imperative that that major languages be represented in the sample. A provisional decision was taken to include 4 of the most commonly-spoken languages: Luo (North), Ateso (North East) and Runyankole/Rukiga (West) and Luganda (Central).[5] Interestingly, according to pupils' responses in the test booklets, 5 major language groups emerged as those most frequently spoken at home. When asked: "What language to you speak at home?" pupils responded as follows:

[5] This decision has not yet been finalized.

Table 1: Languages spoken at home

Language	Number	Percent
Luganda	565	24%
Ateso	509	22%
Langi	294	13%
Runyankole	275	12%
Rukiga	260	11%
Luo	183	8%
Acholi	119	5%
Others	119	5%
Total	2,325	100%

Of course, the purpose of this assessment is to determine the extent to which pupil learning has increased with the new curriculum. This requires a comparison of pupil performance using the "old" curriculum in English and the new one in local languages. However, the language of instruction in a school will be selected not on the basis of a child's mother tongue, but on the basis of the *language of the environment* – that is, the language most widely spoken by children in their day-to-day lives. Based on the figures in the above table, it will be important to confirm the extent to which the 4 languages provisionally selected actually function as the language of the environment for the majority of pupils participating in this MLA.

The design presented in Figure 1 above includes two experimental groups – P2 and P3 – consisting of public (government) schools, and two control groups – again, P2 and P3 – consisting of private schools. The design is considered quasi-experimental because pupils from both experimental and control groups will be pre-tested in 2007 and post-tested in 2008. It is not considered *purely* experimental because selection of pupils and schools was not completely random, due to the importance of stratifying the sample in order to ensure adequate representation of key groups, and due to time and budget constraints influencing the selection.

School attributes

The main criterion for school and pupil selection was to ensure that results could be reported with a 95% confidence interval and a 5% margin of error – standards typically used for student achievement testing. In this case, the populations to which results were to be generalized consisted of *all P2 pupils enrolled in schools in*

the 4 language areas selected. A variety of criteria were used to select schools within each district:

- Location: Urban, peri-urban and rural schools,
- Size: Large and small schools,
- Ownership: Government and private schools,
- Distance: Larger and smaller distances from the district center,
- Boarding type: Day schools, partly boarding and full boarding, and
- Gender: Co-educational, boys only and girls only.

Private schools were selected as possible control schools since many will opt *not* to use English as the medium of instruction with the new curriculum, thereby providing a point of comparison in future MLA iterations with experimental schools that do use local languages. Finally, the most distant schools were excluded due to time and funding limitations.

Once these parameters were established and exclusions were made, remaining schools in the data base were selected randomly by Ministry and UNITY staff.[6]

Sample size

Once a list of eligible schools was generated, the final question concerned sample size. The number of schools and the number of pupils in each school selected was determined in order to minimize sampling error. Based on the total numbers of pupils enrolled in P2 in the 4 language groups cited above, it was calculated that a simple random sample of 400 pupils per group (1,600 total) was required to meet the requirement of a 95% confidence interval and a 5% margin of error (Rosier[7], 1982, p. B-23). However, it is important to note that while the stratification process used in this process tends to give *more* precision than simple random sampling, the two-stage feature of the process tends to give *less* precision (Lohr[8], 1999, p.240). Accordingly, 30 schools for each of the 4 regions (20 experimental and 10 control) and a sub-sample of 20 pupils per school, were included in the sample in order to optimize the sampling error rate. In the end, 2,325 pupils from 117 schools

[6] Further detail on sampling procedures will be provided by members of the Ugandan team.
[7] Rosier, M. (1982) *Sampling and administration manual*. Second IEA Science Study. International Association for the Evaluation of Educational Achievement (IEA).
[8] Lohr, S. L. (1999) *Sampling: Design and Analysis*. Duxbury Press.

participated in this assessment exercise. The total numbers of pupils from the sample attending government and private institutions were as follows:

Table 2: Number of pupils participating in 2007 MLA, by government and private schools

Region	Government	Private	Total
Central	440	160	600
East	399	180	579
North	409	180	589
West	397	160	557
Total	**1,645**	**680**	**2,325**

Test development

From the outset, the UNITY MLA posed two methodological challenges. First, pupil achievement had to be compared from one year to the next using two different curricula: the old, subject-based, traditional one and the new, thematic, pupil-centered one. Second, as noted above, pupil achievement had to be compared from one year to the next using two different languages as the medium of instruction: English with the old curriculum and local languages with the new.

In order to minimize the difficulties arising from these comparisons, an analysis of the old and new curricula was conducted in which common competencies were identified for comparison from the old curriculum to the new. Then, competencies that lent themselves to "linguistically neutral" items were given highest priority. The main consideration here was to minimize linguistic complexities as much as possible, since the same test would be translated into 4 local languages – a change that would certainly bias some items when comparing results from the English test to ones in local languages. For example, in the 2007 test, a pupil might be asked to put "I go" into the present progressive in English - "I am going." In the 2008 test, that item would be translated into 4 different languages, some of which might express the present progressive in more complex or difficult ways than others. Thus, items based on specific vocabulary, grammatical structures or idiomatic expressions could vary in difficulty and therefore be biased toward certain languages. An effort was therefore made to reduce the use of such items or, where possible, to avoid them altogether.

Once common, linguistically neutral competencies had been identified, they were organized into groups in order to ensure they covered the essential parts of the curriculum. Finally, a specifications table or "test blueprint" was developed in order to target different types of thinking skills to be developed for each competency identified. A modified version of Bloom's taxonomy was used for this purpose.

An item writing workshop was then conducted, at which several hundred items were developed by six seasoned P2 teachers, a language expert and a maths expert from the National Curriculum Development Centre's (NCDC), an assessment expert from the Uganda National Examination Board (UNEB), the UNITY Monitoring and Evaluation Specialist, and an international consultant (see *Acknowledgements*). Items were reviewed in Uganda and the US for their psychometric properties, then organized into pilot tests: 2 versions for language, 2 versions for maths and 2 versions of a performance-based test (hereafter called "readiness test").

Once these test "dummies" were developed, test administrators' guides were also developed along with guides for the training of administrators. A Head Teacher's interview instrument was also developed. Eight administrators were trained in the use of these materials, then piloted the tests in the 4 target regions. The completed tests and interview instruments were brought back to Kampala for correction and scoring, after which data were entered. An item analysis was then conducted, taking into consideration item statistics such as item means and item discrimination values, as well as recommendations made by test administrators and project advisors. The best items were selected for the "operational tests" which consisted of one language test (40 items), one maths test (40 items), and one readiness test (6 items). All tests were in English.

For the operational test, 80 Coordinating Center Tutors (CCTs) were recruited to serve as administrators – 20 for each region in 10 teams of 2. Eleven people from MoES headquarters served as supervisors, 1-2 serving in each of the 8 districts. These supervisors were responsible for training the CCTs in test administration, distributing all testing and administration materials, monitoring test administration, and collecting administration reports.

In each school, twenty P2 pupils were chosen randomly by test administrators. Pupils not taking part in the assessment were relocated to another classroom space where they continued normal lessons. At the same time, the test-taking pupils sat in a separate classroom where they first took the language test followed by a 15 minute break, then took the maths test. Afterwards, each pupil was led one by one to another room where he/she performed 6 tasks – 4 for language and 2 for maths. Responses were rated on a 4-point scale with rubrics. Following the administration of pupil tests, the Head Teacher of the school was also interviewed. [9]

Scoring and data entry

After administering the tests, the CCTs returned to their regional centers to submit the booklets and administration reports according to procedures outlined in their MLA administrator's guides. The technical team members then took the test booklets back to Kampala for sorting, tracking each booklet with its own code and ascribing a unique code to each school. Tests were then scored by people working in groups using a common scoring sheet. Next, data were entered by project staff using a template in Excel developed by the international consultant. Data were cleaned by selecting 5 test booklets per district and checking the entered data against the original. An error rate of .03 was found,[10] considered an acceptable level of error to proceed with data analysis. Data sets were sent in electronic format to School-to-School International in the US for analysis and reporting.

Data analysis

Data analysis consisted of three steps: verification of item quality and test reliability, basic descriptive analyses and more advanced inferential analytical procedures, including analysis of variance (ANOVA), post-hoc procedures (Tukey's b, Dunnett's C), Pearson correlations, and Levene's homogeneity of variance test.

[9] Some anomalies were experienced in test administration. For example, in Kabale, some schools were difficult to reach due to difficult terrain and time constraints. At one school, test administrators found primary seven children teaching other classes, with no teachers in place. Only 16 pupils were tested in that school.
[10] This error rate is being confirmed with the Ugandan team.

Data quality

In any test of student achievement, three major sources of error can compromise the quality of the data:

- sampling errors resulting from the sampling design,
- measurement error due to the lack of reliability of the tests or insufficient item discrimination, and
- item bias that favors one type of learner over another (e.g., boys over girls).

This section describes measures taken in these three categories in order to assess item and test quality.

Sampling error

The size of the population and the sampling design yielded the following statistics (see Annex 1 for more details):

- for the language test, the 95% margin of error of the mean was 0.247,[11]
- for the maths test, the 95% margin of error of the mean was 0.180, and
- for the readiness test, the 95% margin of error of the mean was 0.112.

Overall, these margins of error were found to be relatively small, showing very good precision for the estimated means of the three tests. While the margins of error might appear larger for the language and the maths tests, this is normal as values for those tests are greater, with a maximum value of 40, as compared with the readiness test, with a maximum value of 18 (3 points possible for each of 6 items).

Measurement error

In order to measure item quality and test reliability, item means, point-biserial (or "item-to-test") correlations and Cronbach's alpha ("item-to-item") were used. These classical statistics procedures are explained in Bertrand & Blais (2004) and were obtained using SPSS software.

[11] As the sample mean (based on 2325 pupils) for the language test was found to be 19.98 (the minimum value being 0 and the maximum value 40), a 95% margin of error of .247 around 19.98 means that in 95 out of 100 cases, we can be certain that that the mean, based on all P2 pupils, falls within a range of .247 points of the mean – that is, 19.98 - .247 and 19.98 + .247, or 19.73 and 20.23. The scores for other 5% of cases are attributable to chance.

Both language and maths tests consisted of items (i.e., questions or tasks) organized in sets. For example, pupils might be asked to draw a line connecting pictures to words. An example would be given by the administrator, with pupils completing the sample task for that set in their test booklets. Then, a series of 3-4 items would follow in the same format so the pupils could follow the same instructions for all the items in that series.

The advantage of this format is that it reduces the number of different types of instructions pupils must follow in order to complete each item – a strategy often used in contexts where pupils are unfamiliar with testing procedures. However, because items are grouped this way, measuring their reliability individually would be spurious since they are "statistically dependent," meaning that items in a set are related to the same stimulus, so represent the same kind of item. To address this problem, items were grouped into "test-lets" - a technique described in Thissen & Wainer[12] (2001). This grouping resulted in 11 test-lets, or groups of items, in language and 11 test-lets in maths, each test-let consisting of a set of 3-4 items described above. Based on this reorganization, item characteristics were measured using a variety of procedures. A description of these analyses follows.

Language test

Item characteristics on the language and maths tests were measured using three indices: Cronbach alpha, item difficulty, and point biserials. The first of these, the Cronbach alpha coefficient, concerns how well items are correlated with one another. The more consistent the items, the greater the reliability of the test – that is, the more the items all are measuring the same thing, or general construct (e.g. knowledge of how to add double-digit numbers in a maths test). A Cronbach alpha coefficient of 0.7 or higher (up to 1) is generally considered acceptable in student achievement testing. The UNITY MLA 2007 language test obtained a Cronbach's alpha measure of 0.916, indicating an extremely high level of internal consistency and, thus, low measurement error.

[12] Thissen, D., & Wainer, H. (eds) (2001) Test Scoring. Lawrence Erlbaum Associates.

The second of these indices concerns item difficulty. Also called item means or "p-values" when the items are dichotomized (0, 1), this measure simply calculates the proportion of pupils who got an item correct. A p-value of .10 would mean a very difficult item whereas a p-value of .90 indicates a very easy item. In the language test, p-values ranged from .30 to .71 (see table 3 below) – an excellent range for this index.

The third of these indices is the "point-biserial" correlation, or "item-total correlation" – a measure of how well an item discriminates between low and high achievers. An item is said to have good discrimination when students with high exams scores get an item right, and students with low exam scores get an item wrong. The point-biserial correlation is a measure of this relationship – i.e., how well each student performed on each item relative to his/her total exam score. The closer to 1 (the maximum), the greater the discrimination. All the item-total correlations were found very high, meaning that all items are associated with the general language ability measured by the test.

Table 3 presents the results of the 40 language items – again, grouped into 11 test-lets (e.g., "e1t" indicates English test, test-let 1). Two types of information are presented: item means and corrected[13] item-total (point-biserial) correlations. In order to interpret the item means, the values of each item have been standardized to a minimum of 0 and a maximum of 1. Note that the strong item characteristics presented in this table are consistent with the high value of the Cronbach alpha coefficient.

[13] The item-total correlation is said to be "corrected" when the item under analysis is removed from the calculation and is thus not part of the total. This technique helps prevent positive bias in the calculation of the correlation.

Table 3: Language item means and corrected item-total correlation

	Item Means	Corrected Item-Total Correlation
e1t	.71	.673
e2t	.67	.693
e3t	.51	.730
e4t	.52	.828
e5t	.50	.713
e6t	.30	.631
e7t	.43	.682
e8t	.45	.840
e9t	.48	.344
e10t	.60	.451
e11t	.35	.797

Mathematics test

As with the language test, item statistics for the maths test were found to be very strong. The Cronbach alpha coefficient was 0.87 – again, a very high level of internal consistency, well above the .70 threshold, thus indicating very low measurement error. And as can be seen in Table 4 below, item means (p-values) were also good, ranging from .35 to .87, and all item-total correlations were 0.40 and above, a strong level of consistency and thus, low measurement error. Again, these high discrimination values are consistent with the high Cronbach's alpha value. These figures show that the easiest maths item overall was item m3t: "Shade the shape to match the fraction" in which pupils were asked to shade, for example, ½ a triangle when provided an unshaded illustration and the fraction ½. Not surprisingly, the most difficult item was m7t, a sequence of word problems.

Table 4: Mathematics item means and corrected item-total correlation

	Mean	Corrected Item-Total Correlation
m1t	.56	.633
m2t	.54	.630
m3t	.87	.504
m4t	.39	.660
m5t	.38	.682
m6t	.43	.647
m7t	.35	.556
m8t	.70	.461
m9t	.49	.601
m10t	.54	.502
m11t	.40	.560

Readiness test

Finally, on the six items of the readiness test (6 performance-based items: 4 language and 2 maths), the Cronbach's alpha was calculated at 0.79, again, a very high level of internal consistency. Table 5 presents item means and item-test correlations for the readiness test. All 6 items rated pupil performance ability (e.g., "raise your left hand") on a scale of 0-3, where 0 indicated complete inability and 3 indicated complete ability. The easiest readiness item was found to be the first one, p121 ("Stand up") while the hardest one was item p126 ("Write a sentence"). Except for the first item, all corrected item-total correlations (discrimination values) were found to be very good at above 0.50. Item p121 was clearly too easy, with an extremely high mean and low discrimination value of .195.

Table 5: Readiness item means and corrected item-total correlation

	Mean	Corrected Item-Total Correlation
p121	2.85	.195
p122	.88	.636
p123	.72	.599
p124	1.32	.535
p125	.75	.731
p126	.44	.591

Item bias

Our final test of item quality focused on gender bias – that is, did any item show bias for or against boys or girls? To measure this, a statistical procedure called Differential Item Analysis (DIF) was performed on all item test-lets in language and maths – specifically, a technique called "Raju's NCDIF statistic" (Bertrand & Blais, 2004). The Raju's NCDIF procedure found no bias against girls or boys in any of the items. DIF analysis results can be found in Annex 2.

Findings

Findings are organized in three sections: pupil performance, school-level performance and Head Teacher training and views of the new curriculum.

Pupil performance

Pupil performance in language, maths and readiness – the dependent variables – was analyzed in relation to the following independent variables:

- Geography: region, district & zone (urban, peri-urban and rural)
- Home language
- Ownership: public (government) and private
- Age of pupils
- Status as repeaters
- Home environment: books in the home and having a mother who reads, and
- Sex.

Geography

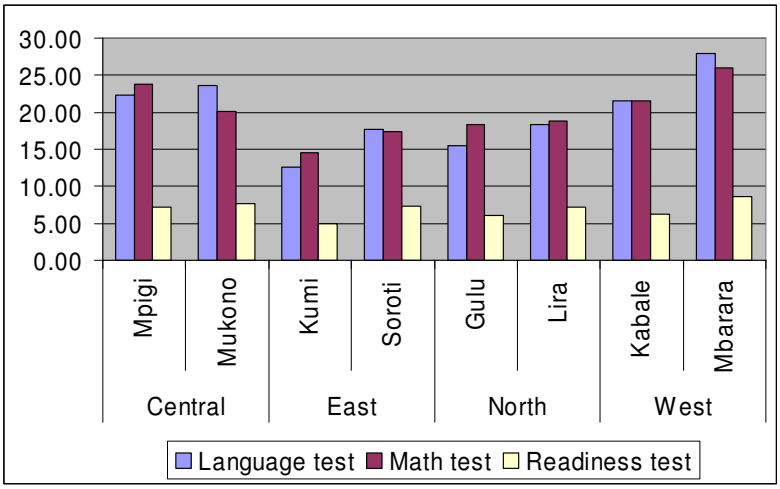
Roughly equal numbers of pupils were tested in all 4 regions, from 557 in the West to 600 in the Central Region. Results varied significantly across regions, with pupils from the West reporting the highest scores of all the regions, followed by those from the Central Region, then those from the Northern Regino. Pupils in the Eastern Region received the lowest scores, as shown in Table 6. All differences between regions were found to be significant (see Annex 3).

Table 6: Mean test scores by region

	Region	Mean	Std. Deviation	N
LanguageTest	Central	23.04	9.638	600
	East	15.27	11.168	579
	North	16.86	11.187	589
	West	24.89	10.900	557
	Total	19.98	11.456	2325
MathTest	Central	21.91	8.088	600
	East	16.01	9.879	579
	North	18.58	9.776	589
	West	23.98	9.007	557
	Total	20.09	9.692	2325
ReadinessTest	Central	7.47	3.590	600
	East	6.20	4.058	579
	North	6.62	4.968	589
	West	7.55	5.016	557
	Total	6.96	4.473	2325

By district, the highest scores were obtained in Mbarara and the lowest in Kumi for all 3 tests as shown in the Figure 2 and the accompanying table:

Figure 2: Results by district
Mean test scores for language and maths tests over 40, readiness over 18

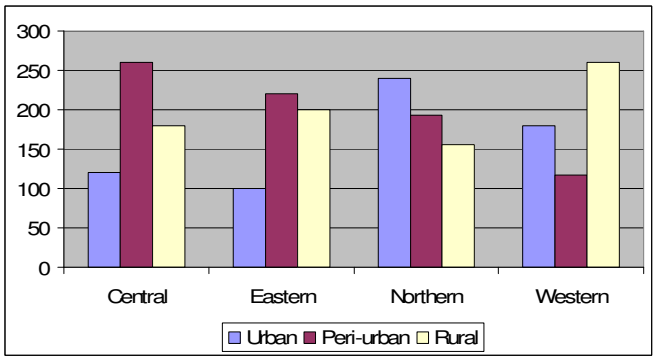


Statistics for Figure 2:

	District	Mean	Std. Deviation	N
LanguageTest	Gulu	15.43	10.042	293
	Kabale	21.51	11.887	260
	Kumi	12.66	10.033	279
	Lira	18.28	12.066	296
	Mbarara	27.85	8.986	297
	Mpigi	22.40	9.799	300
	Mukono	23.68	9.447	300
	Soroti	17.70	11.629	300
	Total	19.98	11.456	2325
MathTest	Gulu	18.34	9.204	293
	Kabale	21.60	9.020	260
	Kumi	14.49	9.616	279
	Lira	18.80	10.321	296
	Mbarara	26.07	8.474	297
	Mpigi	23.75	7.496	300
	Mukono	20.06	8.249	300
	Soroti	17.42	9.926	300
	Total	20.09	9.692	2325
ReadinessTest	Gulu	6.08	4.392	293
	Kabale	6.27	4.216	260
	Kumi	4.94	3.424	279
	Lira	7.16	5.435	296
	Mbarara	8.67	5.385	297
	Mpigi	7.22	3.200	300
	Mukono	7.71	3.931	300
	Soroti	7.37	4.250	300
	Total	6.96	4.473	2325

Proportions of types of schools included in the MLA – urban, peri-urban and rural – varied across regions. Figure 3 and the following table present the number of pupils tested in each type of school by region:

Figure 3: Number of pupils tested by type of school and region



Region	Urban	Peri-urban	Rural	Total
Central	120	260	180	540
Eastern	100	220	200	500
Northern	240	193	156	589
Western	180	117	260	557
Total	640	790	796	2,226

Note that the total number of pupils in the above table is 2,226. This is because some schools could not be categorized as urban, peri-urban or rural.

As might be expected, scores were highest in urban schools, lower in peri-urban schools and lowest in rural ones, as shown in Table 7. Annex 4 also provides additional information, showing that differences in pupil scores were significantly different between all three types of schools for all 3 tests.

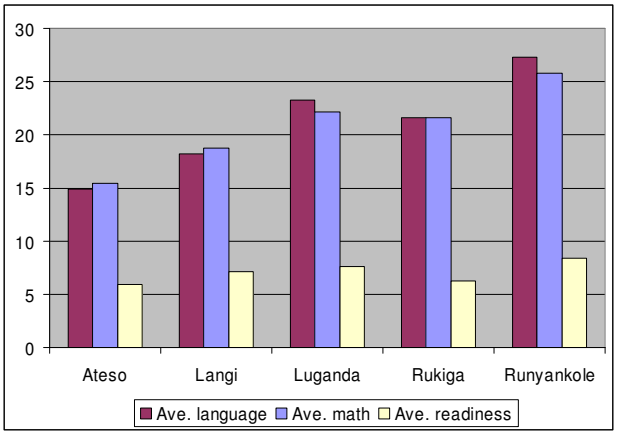
Table 7: Mean test scores by zone: Peri-urban, rural and urban

	Urban	Mean	Std. Deviation	N
LanguageTest	Peri_Urban	20.26	11.430	790
	Rural	14.28	9.116	796
	Urban	27.21	10.085	640
	Total	20.12	11.484	2226
MathTest	Peri_Urban	19.90	9.917	790
	Rural	16.87	8.542	796
	Urban	24.69	8.958	640
	Total	20.19	9.685	2226
ReadinessTest	Peri_Urban	7.04	4.248	790
	Rural	4.67	2.838	796
	Urban	9.79	4.868	640
	Total	6.99	4.501	2226

Home language

When asked to respond to the prompt "My first language is:" pupils reported speaking a total of 27 languages in the home (see list, Annex 5). However, 82% of pupils reported speaking one of five languages: Ateso, Langi, Luganda, Rukiga and Runyankole. Figure 4 and the accompanying table below present results across language groups. As can be seen in Annex 6, the post-hoc test using Dunnett's analysis indicates that for language and math tests, all of these differences are statistically significant, except for the difference between Rukiga and Luganda pupils, which was not significantly different. Results for the performance test were mixed.

Figure 4: Mean test scores by home language
Mean test scores for language and maths tests over 40; readiness over 18



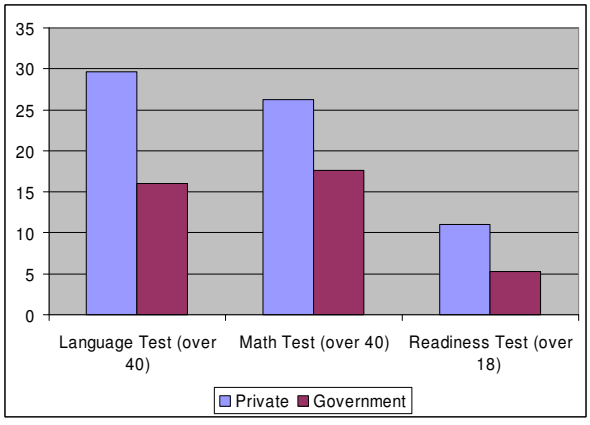
Statistics for Figure 4

	LanguageHome	Mean	Std. Deviation	N
LanguageTest	Ateso	14.88	10.997	509
	Langi	18.17	12.029	294
	Luganda	23.29	9.669	564
	Rukiga	21.59	11.909	260
	Runyankole	27.32	9.059	275
	Total	20.60	11.486	1902
MathTest	Ateso	15.47	9.734	509
	Langi	18.79	10.367	294
	Luganda	22.17	7.915	564
	Rukiga	21.60	9.020	260
	Runyankole	25.83	8.508	275
	Total	20.31	9.708	1902
ReadinessTest	Ateso	5.94	4.007	509
	Langi	7.11	5.424	294
	Luganda	7.60	3.570	564
	Rukiga	6.28	4.222	260
	Runyankole	8.40	5.307	275
	Total	7.02	4.456	1902

Ownership

Approximately 70% of the schools in this sample were government schools and 30% were private ones in each region and district. As the Figure 5 and the accompanying table show, scores for pupils in private schools were substantially higher than in government schools. Annex 7 indicates that these differences are all statistically significant.

Figure 5: Scores by type of school: Private vs. government



Statistics for Figure 5

	Public	Mean	Std. Deviation	N
LanguageTest	Private	29.67	9.175	681
	Public	15.97	9.787	1644
	Total	19.98	11.456	2325
MathTest	Private	26.23	8.533	681
	Public	17.55	8.979	1644
	Total	20.09	9.692	2325
ReadinessTest	Private	11.01	4.569	681
	Public	5.28	3.163	1644
	Total	6.96	4.473	2325

Figures 6, 7, and 8 below show some regional variation between private and government schools, with the greatest disparities being observed in the North and West.

Figure 6: Mean language scores by region: Government & private

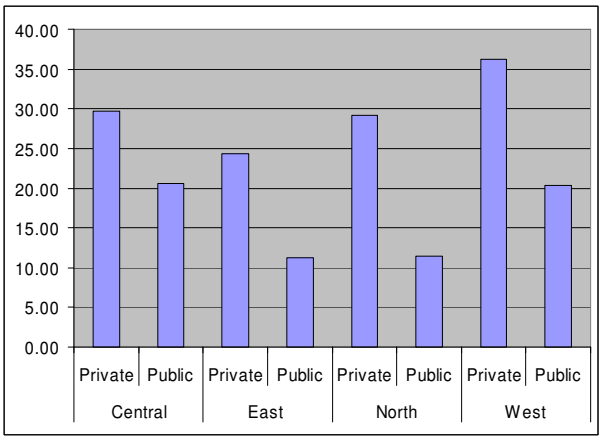


Figure 7: Mean maths scores by region: Government & private

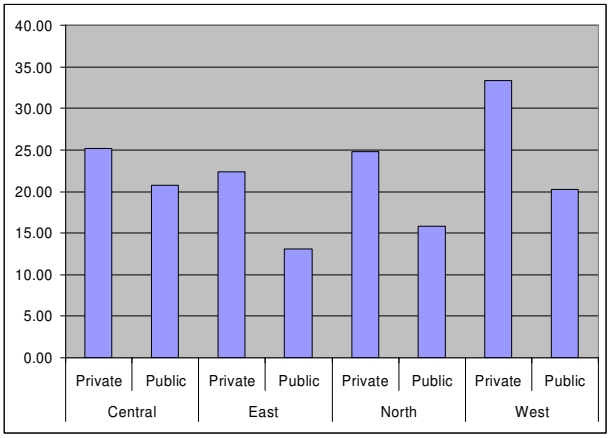
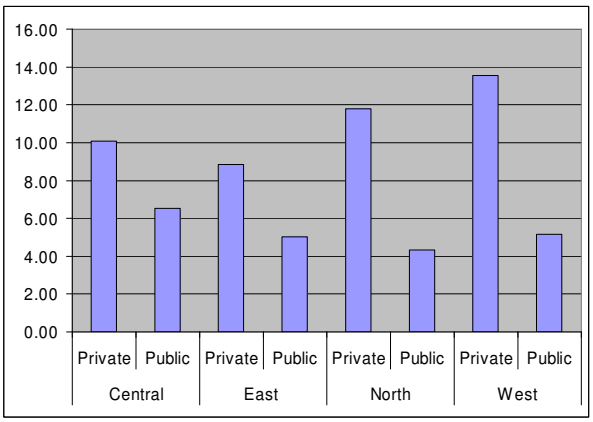


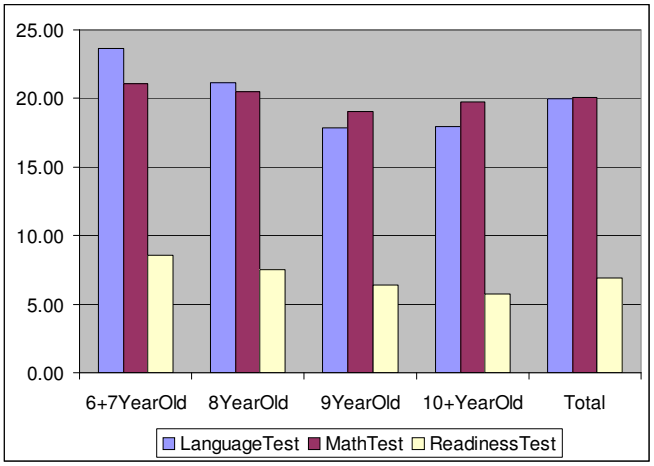
Figure 8: Mean readiness scores by region: Government & private



Age

Pupils reported a range of ages, from 1 (an obvious error) to 15 years old (see Annex 8). Most (86%) fell into the 7-10 age range. Six- and seven-year-olds performed the best, with declining results the following 2 years and leveling off at age 10. Figure 9 and the accompanying table illustrate the rather large differences found in group means for each test.

Figure 9: Mean test scores by age



	Age	Mean	Std. Deviation	N
LanguageTest	6+7YearOld	23.64	11.653	474
	8YearOld	21.12	12.116	630
	9YearOld	17.89	10.933	501
	10+YearOld	17.91	10.329	705
	Total	19.96	11.471	2310
MathTest	6+7YearOld	21.08	9.596	474
	8YearOld	20.49	10.136	630
	9YearOld	19.05	9.317	501
	10+YearOld	19.75	9.556	705
	Total	20.07	9.695	2310
ReadinessTest	6+7YearOld	8.58	4.807	474
	8YearOld	7.52	4.777	630
	9YearOld	6.40	4.000	501
	10+YearOld	5.76	3.845	705
	Total	6.96	4.478	2310

For language and readiness tests, 6- and 7-year-olds performed significantly better than all other age groups. For the maths test, however, the only significant difference was found between 6- and 7-year-olds and 9 year-olds (see post-hoc test comparisons using Dunnett's statistic, Annex 9).

Status as repeaters

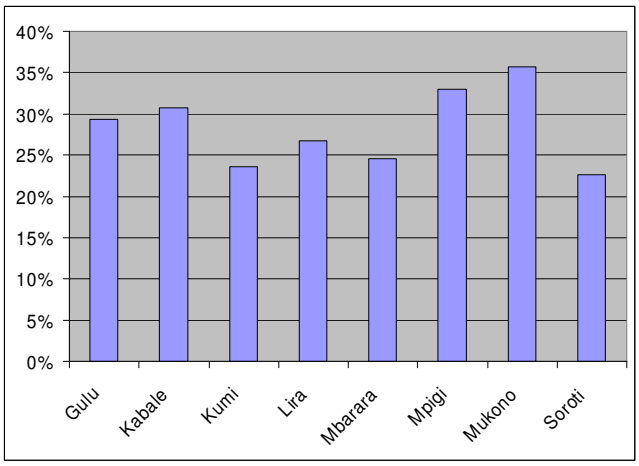
Pupils were asked to circle yes or no in response to the statement "I am repeating this year." Twenty-eight percent of all pupils said yes, with girls and boys reporting nearly exactly the same rates. As shown in Table 8 below, repeaters performed on average 1.5 points lower than nonrepeaters in readiness, 2 points lower in maths, and 3 points lower in language. Annex 11 shows that these mean differences were statistically significant.

It is interesting to note the substantial variation in repetition rates across districts, with the greatest proportion of repeaters coming from Mukono (36%), where language test scores were the second highest (23.68) and maths the third highest (20.06, tied with Mbarara). The lowest proportion was found in Soroti (23%) which ranked 6th in the language test and 7th of 8 districts in the maths test (see Figure 10). Not surprisingly, repetition rates coincided with age, with 7-year-olds reporting lower repetition rates than older P2 pupils (see Figure 11).

Table 8: Mean test scores by status as repeaters

	RepeatYear	Mean	Std. Deviation	N
LanguageTest	Yes	17.84	10.274	658
	No	21.40	11.811	1538
	Total	20.34	11.487	2196
MathTest	Yes	18.87	8.734	658
	No	20.99	10.005	1538
	Total	20.36	9.689	2196
ReadinessTest	Yes	6.01	3.710	658
	No	7.55	4.746	1538
	Total	7.09	4.515	2196

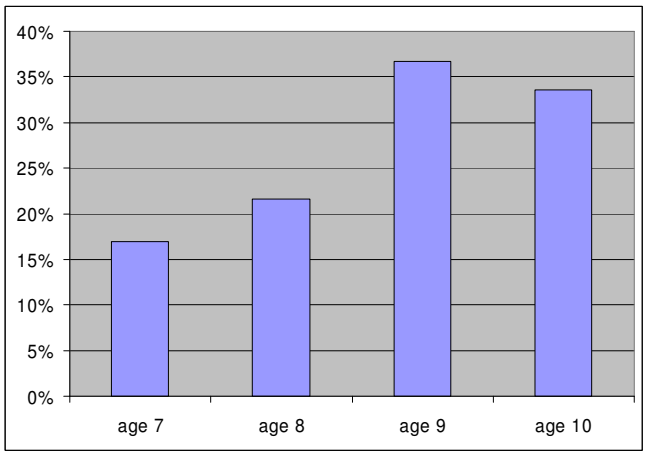
Figure 10: Percentage of pupils reporting as repeaters, by district*



*

Note: 13% (n=40) of pupils reporting from Mbarara did not respond to this question.

Figure 11: Repetition rates by age



Age	Yes	No	Missing	%
6&7	84	365	25	18%
8	136	458	36	22%
9	184	283	34	37%
10	160	296	21	34%

Home environment

In order to have a picture of pupils' home background, they were asked to respond yes or no to two statements: first, "There are books in my home," to which 60% responded yes. To the second statement, "My mother reads at home," 59% said yes. Pupils reporting having books in the home also reported having mothers who read with approximately the same frequency.

Not surprisingly, pupils with books in the home and mothers who read performed approximately 50% better on all 3 tests than ones who responded "no" to both of these questions, as presented in the Tables 9 and 10 below. Annex 11 shows that the differences are statistically significant for all the three tests.

On a regional level, home environment patterns generally reflected test results, with the exception of the number 1 and 2 spots being reversed: pupils from the Northern region, who scored highest of the 4 regions in all 3 tests, had the second highest rates of books in the home and mothers who read, while pupils from the Central region, who scored second highest on all 3 tests, had the highest rates of books and reading mothers (see Figure 12).

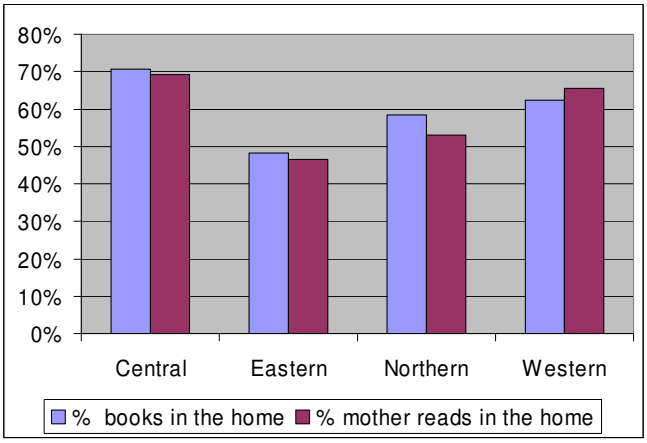
Table 9: Mean test scores by books at home

	BooksHome	Mean	Std. Deviation	N
LanguageTest	Yes	23.60	11.089	1394
	No	14.68	9.929	780
	Total	20.40	11.509	2174
MathTest	Yes	22.87	9.269	1394
	No	16.03	8.831	780
	Total	20.41	9.686	2174
ReadinessTest	Yes	8.15	4.762	1394
	No	5.24	3.273	780
	Total	7.11	4.508	2174

Table 10: Mean test scores by mother who reads

	MotherReads	Mean	Std. Deviation	N
LanguageTest	Yes	23.53	11.035	1363
	No	15.14	10.280	789
	Total	20.45	11.498	2152
MathTest	Yes	22.77	9.174	1363
	No	16.59	9.260	789
	Total	20.50	9.673	2152
ReadinessTest	Yes	8.09	4.669	1363
	No	5.43	3.684	789
	Total	7.12	4.518	2152

Figure 12: Percentage of pupils reporting having books in the home and a mother who reads, by region



Sex

Some of the statistics in this section have already been cited above. However, because of the importance of fostering success for girls in education, this section provides a separate, disaggregated analysis of their scores relative to every independent variable, including some already discussed.

Of the 2,325 P2 pupils tested, 49.5% were girls (n=1,150) and 50.5% boys (n=1,174). As Table 11 shows, girls' and boys' scores were comparable.

Table 11: Mean test scores by sex

	Sex	N	Mean
LanguageTest	Girl	1150	20.17
	Boy	1174	19.82
MathTest	Girl	1150	19.66
	Boy	1174	20.52
ReadinessTest	Girl	1150	6.90
	Boy	1174	7.01

While differences between boys and girls in language and readiness are small, a t-test of independent samples reveals that the difference in maths scores was statistically significant at the .05 probability level (t=-2.139; p=0.033), the difference being in favor of boys (see Annex 13).

On a regional level (Ttable 12), girls' and boys' scores were comparable in the West, Central and East on all 3 tests. However, in the North, boys' scores were significantly higher than girls' in maths. Interestingly, there was virtually no difference in performance between boys and girls on language or readiness items, including maths items in the North.

Table 12: Mean test scores by sex and region

Region		Sex	N	Mean
Central	LanguageTest	Girl	300	23.59
		Boy	300	22.49
	MathTest	Girl	300	21.53
		Boy	300	22.28
	ReadinessTest	Girl	300	7.56
		Boy	300	7.38
East	LanguageTest	Girl	288	15.55
		Boy	291	15.00
	MathTest	Girl	288	15.96
		Boy	291	16.06
	ReadinessTest	Girl	288	6.26
		Boy	291	6.14
North	LanguageTest	Girl	288	16.37
		Boy	300	17.36
	MathTest	Girl	288	17.37
		Boy	300	19.76
	ReadinessTest	Girl	288	6.35
		Boy	300	6.88
West	LanguageTest	Girl	274	25.26
		Boy	283	24.53
	MathTest	Girl	274	23.92
		Boy	283	24.05
	ReadinessTest	Girl	274	7.43
		Boy	283	7.66

A district-level analysis provides more precision, showing that girls and boys performed comparably in all districts except in Gulu, were boys did significantly better than girls in maths. Table 13 provides district-level descriptive data. See also Annex 14 for additional t-test results:

Table 13: Mean test scores by sex and district

District		Sex	N	Mean
Gulu	LanguageTest	Girl	136	14.55
		Boy	156	16.25
	MathTest	Girl	136	16.75
		Boy	156	19.78
	ReadinessTest	Girl	136	5.67
		Boy	156	6.45
Kabale	LanguageTest	Girl	130	22.37
		Boy	130	20.65
	MathTest	Girl	130	22.10
		Boy	130	21.09
	ReadinessTest	Girl	130	6.18
		Boy	130	6.35
Kumi	LanguageTest	Girl	136	12.77
		Boy	143	12.55
	MathTest	Girl	136	14.40
		Boy	143	14.57
	ReadinessTest	Girl	136	4.92
		Boy	143	4.95
Lira	LanguageTest	Girl	152	18.00
		Boy	144	18.57
	MathTest	Girl	152	17.92
		Boy	144	19.74
	ReadinessTest	Girl	152	6.97
		Boy	144	7.35
Mbarara	LanguageTest	Girl	144	27.88
		Boy	153	27.82
	MathTest	Girl	144	25.56
		Boy	153	26.56
	ReadinessTest	Girl	144	8.56
		Boy	153	8.78
Mpigi	LanguageTest	Girl	153	23.33
		Boy	147	21.43
	MathTest	Girl	153	23.34
		Boy	147	24.17
	ReadinessTest	Girl	153	7.24
		Boy	147	7.20
Mukono	LanguageTest	Girl	147	23.86
		Boy	153	23.50
	MathTest	Girl	147	19.64
		Boy	153	20.47
	ReadinessTest	Girl	147	7.89
		Boy	153	7.54
Soroti	LanguageTest	Girl	152	18.03
		Boy	148	17.37
	MathTest	Girl	152	17.35
		Boy	148	17.50
	ReadinessTest	Girl	152	7.45
		Boy	148	7.28

Few differences were found between girls and boys based on their age. The only statistically significant difference was found for the 8 year-olds where boys performed better in the readiness test (see also Annex 14).

Table 14: Mean test scores by sex and age group

Age		Sex	N	Mean
6+7YearOld	LanguageTest	Girl	256	24.45
		Boy	218	22.70
	MathTest	Girl	256	21.36
		Boy	218	20.75
	ReadinessTest	Girl	256	8.96
		Boy	218	8.12
8YearOld	LanguageTest	Girl	360	20.70
		Boy	269	21.74
	MathTest	Girl	360	19.86
		Boy	269	21.36
	ReadinessTest	Girl	360	7.10
		Boy	269	8.10
9YearOld	LanguageTest	Girl	228	17.33
		Boy	273	18.36
	MathTest	Girl	228	18.19
		Boy	273	19.77
	ReadinessTest	Girl	228	6.09
		Boy	273	6.67
10+YearOld	LanguageTest	Girl	297	17.85
		Boy	408	17.96
	MathTest	Girl	297	18.97
		Boy	408	20.32
	ReadinessTest	Girl	297	5.51
		Boy	408	5.94

When analyzed by home language, girls' and boys' performance was comparable in all cases except in the case of maths amongst Acholi speakers, where boys performed significantly better than girls:

Table 15: Mean test scores by sex and language

LanguageHome		Sex	N	Mean
Acholi	LanguageTest	Girl	54	21.22
		Boy	65	22.89
	MathTest	Girl	54	18.83
		Boy	65	22.18
	ReadinessTest	Girl	54	7.61
		Boy	65	9.03
Ateso	LanguageTest	Girl	248	14.98
		Boy	261	14.78
	MathTest	Girl	248	15.28
		Boy	261	15.66
	ReadinessTest	Girl	248	5.99
		Boy	261	5.89
Langi	LanguageTest	Girl	151	17.85
		Boy	143	18.50
	MathTest	Girl	151	17.81
		Boy	143	19.83
	ReadinessTest	Girl	151	6.91
		Boy	143	7.32
Luganda	LanguageTest	Girl	283	23.95
		Boy	281	22.62
	MathTest	Girl	283	21.83
		Boy	281	22.51
	ReadinessTest	Girl	283	7.77
		Boy	281	7.43
Luo	LanguageTest	Girl	87	11.05
		Boy	95	12.33
	MathTest	Girl	87	15.66
		Boy	95	18.25
	ReadinessTest	Girl	87	4.45
		Boy	95	4.85
Rukiga	LanguageTest	Girl	129	22.43
		Boy	131	20.76
	MathTest	Girl	129	22.09
		Boy	131	21.11
	ReadinessTest	Girl	129	6.19
		Boy	131	6.38
Runyankole	LanguageTest	Girl	131	27.24
		Boy	144	27.39
	MathTest	Girl	131	25.34
		Boy	144	26.28
	ReadinessTest	Girl	131	8.27
		Boy	144	8.52

School-level performance

This section presents findings on the statistical significance of relationships between school and teacher characteristics. However, before presenting this information, it is important to note that analyses were limited by several factors.

First, 16 of 117 Head Teacher interviews were missing, thus limiting the confidence of claims made here for the total Head Teacher population in this assessment. Second, the unit of analysis was the school (or Head Teacher); however, only 97 valid matches could be made between the pupil file and the Head Teacher data set. Because those 97 schools were divided into subgroups to perform the analyses, the power of the analyses – i.e., the ability to generalize findings based on a limited sample – was somewhat low. Third, some questions of the Head Teacher schedule had a high rate of non responses, further reducing the power of the statistical analyses. These factors notwithstanding, the following discussion provides information on school characteristics and observations for the consideration of decision-makers.

School characteristics

On average, the 117 schools included in the 2007 MLA had 664 pupils, 339 (51%) of whom were girls. The average teacher population was 17, ranging from 3 at Omoti P/S to 63 at Lira Central; 42% of the teachers in these schools, or 7 of 17 teachers per school on average, were women. On average, 87% of teachers were reported by the Head Teacher to have a formal teaching certificate.

Relationships between schools and pupil achievement

As presented in Table 16 below, the number of male teachers, the number of female teachers, and the total number of teachers in a school are all positively correlated to mean school achievement results. In other words, schools with more teachers, either male or female, tend to get better results in language, mathematics and readiness. Table 16 also shows that the number of teachers with a formal teaching certificate in a school is positively related to school mean achievement. In other words, the more teachers reportedly having a formal teaching certificate in a school, the better the results in all tests.

Table 16: Statistically significant Pearson correlations between Head Teacher variables and mean school achievement results

		Number of male teachers	Number of female teachers	Total number of teachers	Number of certified teachers	Language Test_mean	MathTest_ mean	Readiness Test_mean
Number of male teachers	Pearson Correlation	1	.476**	.882**	.863**	.327**	.295**	.406**
	Sig. (2-tailed)		.000	.000	.000	.001	.003	.000
	N	97	97	97	86	97	97	97
Number of female teachers	Pearson Correlation	.476**	1	.834**	.828**	.263**	.212*	.204*
	Sig. (2-tailed)	.000		.000	.000	.009	.037	.045
	N	97	97	97	86	97	97	97
Total number of teachers	Pearson Correlation	.882**	.834**	1	.982**	.346**	.298**	.364**
	Sig. (2-tailed)	.000	.000		.000	.001	.003	.000
	N	97	97	97	86	97	97	97
Number of certified teachers	Pearson Correlation	.863**	.828**	.982**	1	.271*	.239*	.264*
	Sig. (2-tailed)	.000	.000	.000		.012	.027	.014
	N	86	86	86	86	86	86	86
LanguageTest_mean	Pearson Correlation	.327**	.263**	.346**	.271*	1	.882**	.872**
	Sig. (2-tailed)	.001	.009	.001	.012		.000	.000
	N	97	97	97	86	97	97	97
MathTest_mean	Pearson Correlation	.295**	.212*	.298**	.239*	.882**	1	.786**
	Sig. (2-tailed)	.003	.037	.003	.027	.000		.000
	N	97	97	97	86	97	97	97
ReadinessTest_mean	Pearson Correlation	.406**	.204*	.364**	.264*	.872**	.786**	1
	Sig. (2-tailed)	.000	.045	.000	.014	.000	.000	
	N	97	97	97	86	97	97	97

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

While the preceding table shows how the total number of teachers is positively correlated with school mean results, the next table presents the correlations between the *ratio* of teachers to pupils on the one hand, and mean achievement results in each school, on the other. In this instance, pupil-teacher ratios are an even stronger predictor of success (higher scores) with a correlation around 0.6 than total number of teachers per school, with a correlational measure of about 0.3.

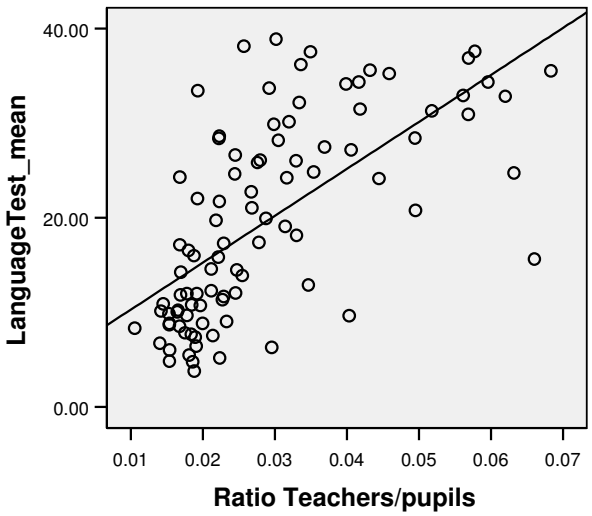
Table 17: Statistically significant Pearson correlations between the ratio of teachers and pupils in a school and the mean school achievement results

		Language Test_mean	MathTest_ mean	Readiness Test_mean	Ratio Teachers/ Pupils
LanguageTest_mean	Pearson Correlation	1	.882**	.872**	.657**
	Sig. (2-tailed)		.000	.000	.000
	N	97	97	97	94
MathTest_mean	Pearson Correlation	.882**	1	.786**	.644**
	Sig. (2-tailed)	.000		.000	.000
	N	97	97	97	94
ReadinessTest_mean	Pearson Correlation	.872**	.786**	1	.630**
	Sig. (2-tailed)	.000	.000		.000
	N	97	97	97	94
Ratio Teachers/Pupils	Pearson Correlation	.657**	.644**	.630**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	94	94	94	94

**. Correlation is significant at the 0.01 level (2-tailed).

The following scatter plot graphically represents this trend toward better performance in relation to lower teacher-pupil ratios. For example, a ratio of 0.05, or 1/20, means 1 teacher to 20 pupils:

Figure 13: Correlation between language test mean and the ratio of the number of teachers to the number of pupils in a school

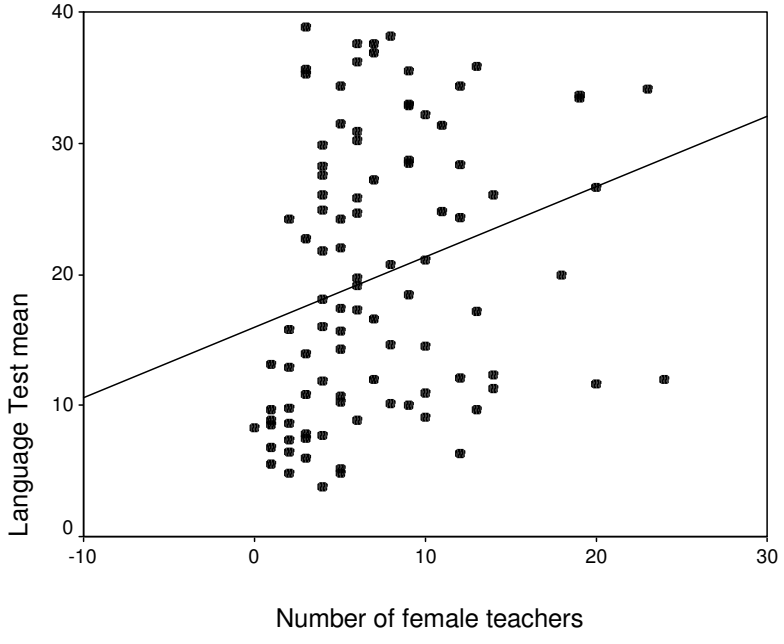


The following scatter plots depict the relation between the total number of *male* teachers in schools and mean pupil scores on the one hand, and the total number of *female* teachers in schools and mean pupil scores, on the other. Both focus on language scores; maths and readiness results exhibited approximately the same

patterns. In essence, they show that the more male teachers there are in a school, the higher the pupil scores are, with a steeper trend line in the case of male teachers. It should be noted that 2 particularly large schools in the sample have over 30 male teachers each. Whereas these 2 schools strengthen the correlations in the case of male teachers, they have less of an effect on the correlation in the case of female teachers.

Figure 14: Correlation between the number of male teachers in school and the school mean result, language test

Figure 15: Correlation between the number of female teachers in school and the school mean result, language test



Head Teacher training and views of the new curriculum

Finally, this section presents findings on characteristics of Head Teachers participating in this MLA, the training they received on the new curriculum, and their views of the new curriculum.

Head Teacher characteristics

As noted above, 101 Head Teachers were interviewed in this MLA. The number of years of experience for Head Teachers was fairly evenly distributed, with 38% reporting up to 5 years of experience, 22% 6-10 years and 38% over 10 years. Most had been in test schools for 5 years or less:

Table 18: Number of years as Head Teacher total and in this school

No. of years	Years total experience	Years as Head Teacher this school
Less than 1	7	22
1 to 5	31	53
6 to 10	22	14
over 10	38	9
no response	2	2
Total	100	100

Fifty-two Head Teachers reported that they had received some kind of training for their role as Head Teachers. Ten of these were provided in the context of the TDMS system and almost all focused on some aspect of management.

Head Teachers' training in the new curriculum[14]

Head Teachers were asked if they had received training in the new curriculum. Most (70 of 100) responded "yes" while 23 said "no," and for 23, no response was recorded. Head Teachers were then asked how long their training had lasted. Durations varied significantly, from 1 to 14 days as follows:

[14] Note that the number of responses for each item varied; analyses were conducted on an item-by-item basis and reflect the number of responses for that item only.

Table 19: Number of days of training in the new curriculum for Head Teachers

Number days training	Number Head Teachers reporting (n=68)
1-2	15
3-6	12
7	33
9-14	8

When asked to rate the quality of their training in the new curriculum, curiously, 72 Head Teachers responded, though only 68 reported having received training. Of the 72 reporting, 54 (75%) rated the training good or excellent:

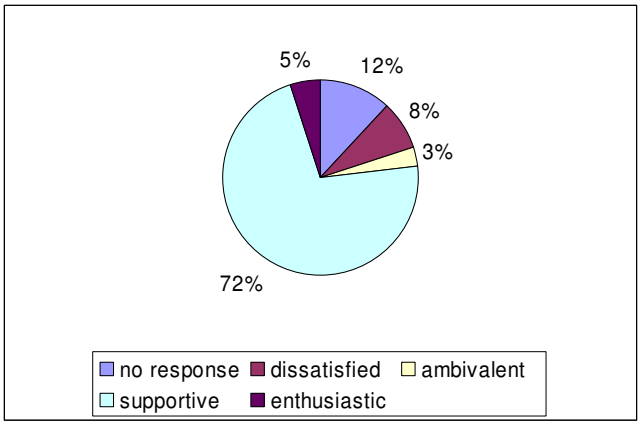
Table 20: Head Teacher rating of the quality of their training in the new curriculum

Quality of training	No.	%
poor	8	11%
adequate	10	14%
good	50	69%
excellent	4	6%
Total	72	100%

Head Teachers' views of the new curriculum

Whether they had received training or not, the vast majority (72%) of Head Teachers said they were supportive of the new curriculum; however, only 5% said they were "enthusiastic" (see Figure 16 below).

Figure 16: Head Teacher attitudes toward the new curriculum
N=100



The number of days of training Head Teachers received on the new curriculum was strongly correlated with the quality of the training as well as their views on the new curriculum. Interestingly, most Head Teachers who were supportive of, or enthusiastic about, the new curriculum had received 1-7 days of training, the majority of those having received 5-7 days of training (see Figures 17 and 18). This finding suggests that there might be limited value in providing training beyond 7 days.

Figure 17: Views of Head Teachers concerning the quality of their training on the new curriculum, by number of days of training.
Number of responses.

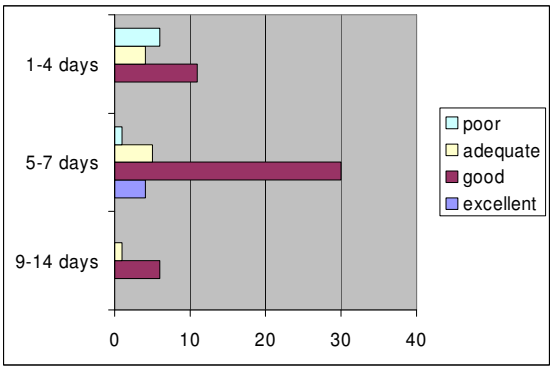
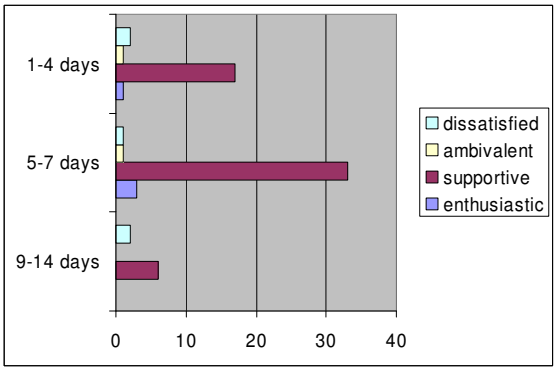


Figure 18: Views of Head Teachers concerning the new curriculum, by number of days of training.
Number of responses.



Discussion

It is important to remember that the 2007 MLA exercise was a "snapshot," a first glance at the profile of the population of pupils to measured over time in order to answer one overarching question: "Are pupils learning more with the new curriculum?" As such, results from this one-time measure must be viewed with caution, understanding that though the sampling design was sufficiently powerful and the items statistics were extremely good, it is nevertheless a single measure and, as such, could mask realities that could not be captured in this test administration. A more reliable picture can only emerge over time, when different groups of randomly-sampled pupils take part in the same exercise and patterns can be identified with greater confidence. However, with a sample size of 2,325 pupils and strong item statistics recorded in this exercise, we are confident in the

ability of this exercise to reliably measure improvement of pupil scores over the next two years.

It is also important to remember that this being the first administration of this MLA, many lessons were learned concerning test construction, administration and scoring procedures as well as sampling considerations. All of these lessons will be incorporated into future iterations of the MLA under the UNITY Project.

In particular, during the piloting of these tests, it was found that pupils experienced substantial problems simply in following directions. One reason for this difficulty appeared to be the question of language: according to administrators' reports, pupils often had few or no English skills, especially in rural areas, and thus could not often understand the items. In the operational test, administrators provided instructions in local languages for each item set, but all items were printed in English, and only English responses could be considered correct. Since this assessment did not include a separate feature to assess pupil mastery of English or the amount of English used by teachers in the average classroom, it is impossible to determine the extent to which pupil performance was influenced by their lack of mastery of English, and their limited English no doubt reduced their chances of success.

These things being said, a number of observations can be made in light of the 2007 MLA exercise. One of the most striking findings from the 2007 MLA exercise concerns results observed by gender. It is perhaps surprising, but certainly reassuring, that of all 3 tests, girls' performance was only found to be significantly poorer than boys in maths and this, only in Gulu district and in the Acholi language.

Another striking finding concerns differences observed on the basis of geography (region district, urban/peri-urban, rural), school ownership (government vs. private), and school conditions (qualification of teachers, PTRs), all of which yielded significant differences across the assessment's population. What emerges is a set of distinct patterns by subgroup - region, district, school type and zone – that differ significantly from one group to the next.

Yet while significant differences were found between these groups, less pronounced were differences *within* subgroups – for example, children in rural government schools in a given district. In other words, it appears that a substantial number of disadvantaged groups exist, that their identity is fairly consistent, and results are relatively uniform within those groups.

In short, the results of the 2007 MLA paint two distinct pictures: higher-achieving pupils that correlate strongly with certain characteristics, and lower-achieving pupils who correlate with other characteristics. In this MLA, higher-achieving P2 pupils:

- attended an urban school,
- attended a private school,
- came from the Western region,
- spoke Runyankole at home,
- were 6-7 years old,
- were nonrepeaters,
- had books in home and mother who reads,
- had teachers with formal certification,
- had a large number of teachers in their school, especially male ones, and
- came from classes with relatively low pupil-teacher ratios (i.e., small classes).

Conversely, lower achieving P2 pupils:

- attended a peri-urban or rural school,
- attended a government school,
- came from the Eastern region,
- spoke Ateso at home,
- were 9-10 years old,
- were repeating P2,
- had no books in home and their mother did not read,
- teachers had no formal certification,
- had a smaller number of teachers in their school,
- had a larger proportion of female teachers in their school, and
- came from classes with relatively high pupil-teacher ratios (i.e., large classes).

Future iterations of the MLA will show whether these characteristics are constant and equally pronounced. Moreover, the anomalies cited in this assessment will certainly warrant close scrutiny in future MLA exercises. For example, the variations in the relationship between repetition and pupil performance, girls' performance on maths in selected districts, results of low achievers, the mean difference in results between regions or districts, and the improvement of low achievers over time, will all merit continued attention to determine whether these patterns are consistent and, if so, why. Results of these observations might lead to the conclusion that attention must be paid to these subpopulations and their particular circumstances that, depending on their scale, might present an opportunity to help teachers and pupils in a state particular need with assistance that can bring about a significant improvement on a targeted basis. Ideas for such opportunities are offered in the recommendations section that follows.

Recommendations

Methodological considerations

In future iterations of the MLA, the following recommendations should be taken into account:

1. Use tests and items in the forthcoming 2008 and 2009 MLA exercises that are matched to the 2007 tests – that is, do not simply use translations of the 2007 test in 2008 and 2009, which would generate a temptation to copy and "teach to the test." Rather, generate comparable items and pilot them to ensure similar item characteristics to 2007.
2. Ensure careful language selection and translation in future MLA exercises.
    a. Select the main languages for 2008 and 2009 iterations of the MLA on the basis of the languages most widely spoken as the language of the environment by the majority of pupils participating in the MLA. The 5 languages identified as the most frequently spoken ones in this report would be the most likely choices.
    b. Conduct back-to-back translation of the tests – e.g., translating test items and instructions from English into Ateso, then from Ateso back into English. This method ensures that translations are as simple, clear and free of linguistic errors as possible.
3. Revise interview instruments to include key information and categories.

a. Ensure the inclusion of data on the sex of teachers of pupils being tested. It should be noted that the Head Teacher interview instrument did not include a provision for stating the sex teacher responsible for the MLA class in the 2007 exercise. However, the 2008 MLA will include interviews with teachers whose pupils are participating in the test, and will include this item.
b. To facilitate analysis, develop categories for the number of days of training in new curriculum received by teachers and Head Teachers – e.g., 1-4 days, 5-7 days, etc.
c. To facilitate analysis, develop sets of categories for teachers and Head Teachers by level formal training received (e.g., Grade III).

4. Revise selected items:
    a. Similar items: Responses to some items were remarkable similar – e.g., questions concerning books in the home and having a mother who reads. In such cases, consider separating the items so pupils' understanding of one does not influence their response to another.
    b. Easy items: Items with particularly high p-scores (e.g., p121) should be revised in order to improve their discrimination power.
5. Improve training and monitoring processes.
    a. Provide additional training of test administrators and monitoring of test administration and interviews in the field. Training should include a discussion of the importance of interviewing everyone in an administrator's zone, and recording a response for each item on the interview instrument.
    b. Provide additional training for and monitoring of data entry technicians. Training should include monitoring of data entry and data cleaning as data are being entered.
    c. Ensure that pupils' names are entered in the data files.

Programmatic recommendations

1. Compare results of this assessment to ones conducted by UNEB and others in order to validate findings and identify differences of method or results. Where differences are found, investigate the details and derive lessons from each study in order to improve the reliability of data collection in future MLA iterations.

2. If characteristics of low-performing pupils remain constant or anomalies are repeated, examine factors associated with these phenomena and explore intervention modalities that might be able to provide assistance to the most vulnerable subgroups.
3. Based on patterns emerging from findings of this and future MLA exercises, identify curricular objectives with which pupils have the greatest difficulties, and provide recommendations concerning the nature of their difficulties and possible courses of remediation.
4. Based on perceptions of Head Teachers and teachers concerning their training in the new curriculum, make possible adjustments to training content or duration in future programs.

References

Bertrand, R. & Valiquette, C. (1986) *Pratique de l'analyse statistique des données*. Québec: Presses de l'Université du Québec.

Bertrand, R. & Blais, J-G. (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items*. Québec: Presses de l'Université du Québec.

Cochran, W. G. (1977). *Sampling techniques*. New York : John Wiley.

Glass, G. V., & Hopkins, K. D. (1996) *Statistical methods in education and psychology*. Boston: Allyn & Bacon.

Lohr, S. L. (1999) *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.

Raju, N.S, van der Linden, W.J., & Fleer, P.F. (1995) IRT-based internal measure of differential functioning of items and tests. *Applied Psychological Measurement, 19*, 4, 353-368.

Rosier, M. (1982) Sampling and administration manual. Second IEA Science Study: International Association for the Evaluation of Educational Achievement (IEA).

Thissen, D., & Wainer. H. (eds). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.

Annex 1: Computation of the 95% margin of error

Following Rosier (1982, p.B-11) the intraclass correlation coefficient (rho) can be computed by the formula, for each stratum and each test:

$$rho = [(n^* \times BSV) - BPV] / [(n^* - 1) \times BPV]$$

where n^* is the average number of pupils by school in the stratum,

BSV is the between schools variance and

BPV is the between pupils variance.

Then (Snedecor & Cochran, 1971, p.328, p.589) the variance of the mean VAR_H for each stratum and each test would be

$$VAR\_H = [(rho \times BPV) / n_S] + [((1-rho) \times BPV) / n_P]$$

where n_S is the number of schools in the stratum and

n_P is the number of pupils in the stratum.

The final variance of the mean for all strata is then (Lohr, 1999, p.100) the sum of the VAR_H weighted by a certain factor, that is

$$VAR\_ALL = sum [VAR\_H \times (1 - n_H/N_H) \times (1/n_H) \times (N_H/N)^2]$$

where n_H is the number of schools sampled in stratum H and

N_H is the total number of schools in the stratum H.

And the standard error of the mean (se (mean)) is the square root of the VAR_ALL.

Finally the 95% margin of error (supposing normal distribution) is 1,96 times this standard error.

Given those formulas, we computed that:

- for the language test, the 95% margin of error is 0,247;
- for the maths test, the 95% margin of error is 0,180;
- for the readiness test, the 95% margin of error is 0,112.

Annex 2: DIF analyses using Raju's NCDIF statistic for polytomous data

LANGUAGE TEST

Item number	NCDIF value	Critical value
1	0.006046	0.1
2	0.003794	0.1
3	0.00084	0.1
4	0.003842	0.1
5	0.002964	0.1
6	0.002085	0.1
7	0.001107	0.1
8	0.003213	0.1
9	0.00105	0.006
10	0.004202	0.054
11	0.004246	0.1

None of the observed NCDIF value exceeds the critical value indicating that no language item was found DIF.

MATHS TEST

Item number	NCDIF value	Critical value
1	0.003856	0.1
2	0.001908	0.054
3	0.000082	0.054
4	0.003907	0.1
5	0.004318	0.1
6	0.003384	0.1
7	0.002574	0.054
8	0.000507	0.054
9	0.018728	0.1
10	0.025722	0.1
11	0.004913	0.1

NB: None of the observed NCDIF value exceeds the critical value indicating that no maths item was found DIF.

Annex 3: Multiple comparisons of test means using Dunnett's test, by region

Dependent Variable		(I) Region	(J) Region	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
LanguageTest	Dunnett C	Central	East	7,77*	,608	6,20	9,33
			North	6,18*	,606	4,62	7,74
			West	-1,85*	,607	-3,42	-,29
		East	Central	-7,77*	,608	-9,33	-6,20
			North	-1,59	,654	-3,27	,10
			West	-9,62*	,655	-11,30	-7,93
		North	Central	-6,18*	,606	-7,74	-4,62
			East	1,59	,654	-,10	3,27
			West	-8,03*	,653	-9,71	-6,35
		West	Central	1,85*	,607	,29	3,42
			East	9,62*	,655	7,93	11,30
			North	8,03*	,653	6,35	9,71
MathTest	Dunnett C	Central	East	5,89*	,527	4,54	7,25
			North	3,33*	,521	1,99	4,67
			West	-2,08*	,505	-3,38	-,78
		East	Central	-5,89*	,527	-7,25	-4,54
			North	-2,57*	,575	-4,05	-1,08
			West	-7,97*	,561	-9,42	-6,53
		North	Central	-3,33*	,521	-4,67	-1,99
			East	2,57*	,575	1,08	4,05
			West	-5,41*	,555	-6,84	-3,98
		West	Central	2,08*	,505	,78	3,38
			East	7,97*	,561	6,53	9,42
			North	5,41*	,555	3,98	6,84
ReadinessTest	Dunnett C	Central	East	1,27*	,223	,69	1,85
			North	,85*	,252	,20	1,49
			West	-,08	,258	-,75	,58
		East	Central	-1,27*	,223	-1,85	-,69
			North	-,42	,265	-1,11	,26
			West	-1,35*	,271	-2,05	-,66
		North	Central	-,85*	,252	-1,49	-,20
			East	,42	,265	-,26	1,11
			West	-,93*	,295	-1,69	-,17
		West	Central	,08	,258	-,58	,75
			East	1,35*	,271	,66	2,05
			North	,93*	,295	,17	1,69

Based on observed means.

*. The mean difference is significant at the ,05 level.

Annex 4: Multiple comparisons using Dunnett's test by zone: urban, peri-urban, and rural

Dependent Variable		(I) Urban	(J) Urban	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
LanguageTest	Dunnett C	Peri_Urban	Rural	5,76*	,523	4,53	6,99
			Urban	-7,40*	,574	-8,74	-6,05
		Rural	Peri_Urban	-5,76*	,523	-6,99	-4,53
			Urban	-13,16*	,515	-14,37	-11,95
		Urban	Peri_Urban	7,40*	,574	6,05	8,74
			Rural	13,16*	,515	11,95	14,37
MathTest	Dunnett C	Peri_Urban	Rural	3,08*	,471	1,97	4,18
			Urban	-4,91*	,507	-6,11	-3,72
		Rural	Peri_Urban	-3,08*	,471	-4,18	-1,97
			Urban	-7,99*	,467	-9,09	-6,89
		Urban	Peri_Urban	4,91*	,507	3,72	6,11
			Rural	7,99*	,467	6,89	9,09
ReadinessTest	Dunnett C	Peri_Urban	Rural	2,36*	,183	1,93	2,79
			Urban	-2,81*	,250	-3,40	-2,23
		Rural	Peri_Urban	-2,36*	,183	-2,79	-1,93
			Urban	-5,17*	,221	-5,69	-4,65
		Urban	Peri_Urban	2,81*	,250	2,23	3,40
			Rural	5,17*	,221	4,65	5,69

Based on observed means.

*. The mean difference is significant at the ,05 level.

Annex 5: Language spoken at home

Language	Number pupils	Percent
Acholi	119	5.12%
Arabic	1	0.04%
Ateso	509	21.89%
English	4	0.17%
Hindu	1	0.04%
Japadhola	3	0.13%
Kikuyu	1	0.04%
Kiswahili	7	0.30%
Kumam	59	2.54%
Kupsabin	1	0.04%
Langi	294	12.65%
Luganda	565	24.30%
Lugbara	5	0.22%
Lugisu	3	0.13%
Lugwere	4	0.17%
Lukhonzo	1	0.04%
Lumasaba	1	0.04%
Luo	183	7.87%
Lusamya	1	0.04%
Lusoga	15	0.65%
Madi	1	0.04%
Ngakarimojony	1	0.04%
Rukiga	260	11.18%
Runyankole	275	11.83%
Runyarwanda	4	0.17%
Rutoro	1	0.04%
Swahili	5	0.22%
Total	2325	100.00%

Annex 6: Multiple comparisons using Dunnett's test by language group

Dependent Variable		(I) LanguageHome	(J) LanguageHome	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
LanguageTest	Dunnett C	Ateso	Langi	-3,29*	,854	-5,63	-,95
			Luganda	-8,41*	,635	-10,15	-6,67
			Rukiga	-6,71*	,885	-9,14	-4,28
			Runyanko	-12,44*	,732	-14,45	-10,43
		Langi	Ateso	3,29*	,854	,95	5,63
			Luganda	-5,12*	,811	-7,34	-2,89
			Rukiga	-3,42*	1,019	-6,22	-,62
			Runyanko	-9,15*	,889	-11,59	-6,71
		Luganda	Ateso	8,41*	,635	6,67	10,15
			Langi	5,12*	,811	2,89	7,34
			Rukiga	1,70	,843	-,62	4,01
			Runyanko	-4,03*	,681	-5,90	-2,16
		Rukiga	Ateso	6,71*	,885	4,28	9,14
			Langi	3,42*	1,019	,62	6,22
			Luganda	-1,70	,843	-4,01	,62
			Runyanko	-5,73*	,919	-8,25	-3,21
		Runyanko	Ateso	12,44*	,732	10,43	14,45
			Langi	9,15*	,889	6,71	11,59
			Luganda	4,03*	,681	2,16	5,90
			Rukiga	5,73*	,919	3,21	8,25
MathTest	Dunnett C	Ateso	Langi	-3,32*	,743	-5,35	-1,28
			Luganda	-6,70*	,545	-8,19	-5,21
			Rukiga	-6,12*	,706	-8,06	-4,19
			Runyanko	-10,36*	,670	-12,20	-8,52
		Langi	Ateso	3,32*	,743	1,28	5,35
			Luganda	-3,38*	,690	-5,28	-1,49
			Rukiga	-2,81*	,824	-5,07	-,55
			Runyanko	-7,04*	,793	-9,22	-4,87
		Luganda	Ateso	6,70*	,545	5,21	8,19
			Langi	3,38*	,690	1,49	5,28
			Rukiga	,58	,651	-1,21	2,36
			Runyanko	-3,66*	,612	-5,34	-1,98
		Rukiga	Ateso	6,12*	,706	4,19	8,06
			Langi	2,81*	,824	,55	5,07
			Luganda	-,58	,651	-2,36	1,21
			Runyanko	-4,24*	,759	-6,32	-2,15
		Runyanko	Ateso	10,36*	,670	8,52	12,20
			Langi	7,04*	,793	4,87	9,22
			Luganda	3,66*	,612	1,98	5,34
			Rukiga	4,24*	,759	2,15	6,32
ReadinessTest	Dunnett C	Ateso	Langi	-1,18*	,363	-2,17	-,18
			Luganda	-1,66*	,233	-2,30	-1,02
			Rukiga	-,35	,316	-1,22	,52
			Runyanko	-2,47*	,366	-3,47	-1,46
		Langi	Ateso	1,18*	,363	,18	2,17
			Luganda	-,49	,350	-1,45	,48
			Rukiga	,83	,411	-,30	1,96
			Runyanko	-1,29*	,450	-2,53	-,06
		Luganda	Ateso	1,66*	,233	1,02	2,30
			Langi	,49	,350	-,48	1,45
			Rukiga	1,31*	,302	,48	2,14
			Runyanko	-,81	,354	-1,78	,16
		Rukiga	Ateso	,35	,316	-,52	1,22
			Langi	-,83	,411	-1,96	,30
			Luganda	-1,31*	,302	-2,14	-,48
			Runyanko	-2,12*	,414	-3,25	-,98
		Runyanko	Ateso	2,47*	,366	1,46	3,47
			Langi	1,29*	,450	,06	2,53
			Luganda	,81	,354	-,16	1,78
			Rukiga	2,12*	,414	,98	3,25

Based on observed means.

*. The mean difference is significant at the ,05 level.

Annex 7: T-Test by ownership, private vs. government (public)
Independent samples test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
LanguageTest	Equal variances assumed	13,418	,000	31,260	2323	,000
	Equal variances not assumed			32,101	1344,357	,000
MathTest	Equal variances assumed	4,634	,031	21,578	2323	,000
	Equal variances not assumed			22,044	1328,544	,000
ReadinessTest	Equal variances assumed	197,482	,000	34,780	2323	,000
	Equal variances not assumed			30,031	959,821	,000

Annex 8: Self-reported ages of pupils

Age	Frequency	Percent
1	1	0.0%
2	4	0.2%
4	5	0.2%
5	5	0.2%
6	85	3.7%
7	389	16.7%
8	630	27.1%
9	501	21.5%
10	477	20.5%
11	143	6.2%
12	63	2.7%
13	14	0.6%
14	7	0.3%
15	1	0.0%
Total	2325	100.0%

Annex 9: Multiple comparisons using Dunnett's test by age group

Dependent Variable		(I) Age	(J) Age	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
						Lower Bound	Upper Bound
LanguageTest	Dunnett C	6+7YearOld	8YearOld	2,52*	,721	,67	4,38
			9YearOld	5,75*	,725	3,88	7,62
			10+YearOld	5,73*	,662	4,02	7,43
		8YearOld	6+7YearOld	-2,52*	,721	-4,38	-,67
			9YearOld	3,23*	,687	1,46	5,00
			10+YearOld	3,21*	,620	1,61	4,80
		9YearOld	6+7YearOld	-5,75*	,725	-7,62	-3,88
			8YearOld	-3,23*	,687	-5,00	-1,46
			10+YearOld	-,02	,624	-1,63	1,59
		10+YearOld	6+7YearOld	-5,73*	,662	-7,43	-4,02
			8YearOld	-3,21*	,620	-4,80	-1,61
			9YearOld	,02	,624	-1,59	1,63
MathTest	Dunnett C	6+7YearOld	8YearOld	,59	,598	-,95	2,13
			9YearOld	2,02*	,606	,46	3,59
			10+YearOld	1,32	,569	-,14	2,79
		8YearOld	6+7YearOld	-,59	,598	-2,13	,95
			9YearOld	1,43	,580	-,06	2,93
			10+YearOld	,73	,541	-,66	2,13
		9YearOld	6+7YearOld	-2,02*	,606	-3,59	-,46
			8YearOld	-1,43	,580	-2,93	,06
			10+YearOld	-,70	,550	-2,12	,72
		10+YearOld	6+7YearOld	-1,32	,569	-2,79	,14
			8YearOld	-,73	,541	-2,13	,66
			9YearOld	,70	,550	-,72	2,12
ReadinessTest	Dunnett C	6+7YearOld	8YearOld	1,05*	,292	,30	1,80
			9YearOld	2,17*	,284	1,44	2,91
			10+YearOld	2,82*	,264	2,14	3,50
		8YearOld	6+7YearOld	-1,05*	,292	-1,80	-,30
			9YearOld	1,12*	,261	,45	1,79
			10+YearOld	1,77*	,239	1,15	2,38
		9YearOld	6+7YearOld	-2,17*	,284	-2,91	-1,44
			8YearOld	-1,12*	,261	-1,79	-,45
			10+YearOld	,65*	,230	,05	1,24
		10+YearOld	6+7YearOld	-2,82*	,264	-3,50	-2,14
			8YearOld	-1,77*	,239	-2,38	-1,15
			9YearOld	-,65*	,230	-1,24	-,05

Based on observed means.

*. The mean difference is significant at the ,05 level.

Annex 10: T-Test by repeating year (yes vs no)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
LanguageTest	Equal variances assumed	42,698	,000	-6,720	2194	,000
	Equal variances not assumed			-7,104	1416,441	,000
MathTest	Equal variances assumed	21,575	,000	-4,731	2194	,000
	Equal variances not assumed			-4,995	1411,557	,000
ReadinessTest	Equal variances assumed	102,654	,000	-7,374	2194	,000
	Equal variances not assumed			-8,126	1570,042	,000

Annex 11: T-Test for books at home and mother reading at home

Independent samples test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
LanguageTest	Equal variances assumed	38,763	,000	18,656	2172	,000
	Equal variances not assumed			19,245	1764,914	,000
MathTest	Equal variances assumed	2,420	,120	16,795	2172	,000
	Equal variances not assumed			17,025	1678,571	,000
ReadinessTest	Equal variances assumed	233,484	,000	15,176	2172	,000
	Equal variances not assumed			16,800	2083,113	,000

Annex 12: T-Test comparing boys' and girls' performance

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
LanguageTest	Equal variances assumed	2,741	,098	,738	2322	,460
	Equal variances not assumed			,738	2316,503	,461
MathTest	Equal variances assumed	,297	,586	-2,139	2322	,033
	Equal variances not assumed			-2,139	2320,892	,033
ReadinessTest	Equal variances assumed	,516	,473	-,603	2322	,547
	Equal variances not assumed			-,603	2321,089	,547

Annex 13: T-test by sex vs. other variables

Grouping variable: age

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means		
Age			F	Sig.	t	df	Sig. (2-tailed)
6+7YearOld	LanguageTest	Equal variances assumed	,654	,419	1,634	472	,103
		Equal variances not assumed			1,639	464,688	,102
	MathTest	Equal variances assumed	,436	,509	,691	472	,490
		Equal variances not assumed			,692	463,045	,489
	ReadinessTest	Equal variances assumed	4,422	,036	1,914	472	,056
		Equal variances not assumed			1,926	468,840	,055
8YearOld	LanguageTest	Equal variances assumed	1,022	,312	-1,072	627	,284
		Equal variances not assumed			-1,068	569,568	,286
	MathTest	Equal variances assumed	1,154	,283	-1,834	627	,067
		Equal variances not assumed			-1,821	561,290	,069
	ReadinessTest	Equal variances assumed	4,248	,040	-2,609	627	,009
		Equal variances not assumed			-2,581	552,518	,010
9YearOld	LanguageTest	Equal variances assumed	,141	,708	-1,046	499	,296
		Equal variances not assumed			-1,047	484,934	,296
	MathTest	Equal variances assumed	3,390	,066	-1,895	499	,059
		Equal variances not assumed			-1,882	468,750	,060
	ReadinessTest	Equal variances assumed	8,318	,004	-1,616	499	,107
		Equal variances not assumed			-1,633	497,071	,103
10+YearOld	LanguageTest	Equal variances assumed	,021	,886	-,146	703	,884
		Equal variances not assumed			-,145	635,376	,884
	MathTest	Equal variances assumed	1,507	,220	-1,853	703	,064
		Equal variances not assumed			-1,868	656,212	,062
	ReadinessTest	Equal variances assumed	3,527	,061	-1,448	703	,148
		Equal variances not assumed			-1,466	663,544	,143

Annex 13 (cont'd): T-test by sex vs other variables

Grouping variable: language at home

Independent Samples Test

LanguageHome			Levene's Test for Equality of Variances: F	Levene's Test for Equality of Variances: Sig.	t-test for Equality of Means: t	t-test for Equality of Means: df	t-test for Equality of Means: Sig. (2-tailed)
Acholi	LanguageTest	Equal variances assumed	1,762	,187	-,943	117	,347
		Equal variances not assumed			-,935	108,157	,352
	MathTest	Equal variances assumed	1,164	,283	-2,185	117	,031
		Equal variances not assumed			-2,170	109,608	,032
	ReadinessTest	Equal variances assumed	,008	,929	-1,564	117	,120
		Equal variances not assumed			-1,557	110,844	,122
Ateso	LanguageTest	Equal variances assumed	1,475	,225	,207	507	,836
		Equal variances not assumed			,207	502,362	,836
	MathTest	Equal variances assumed	,024	,876	-,436	507	,663
		Equal variances not assumed			-,436	504,782	,663
	ReadinessTest	Equal variances assumed	,187	,665	,301	507	,764
		Equal variances not assumed			,300	501,127	,764
Langi	LanguageTest	Equal variances assumed	,315	,575	-,462	292	,645
		Equal variances not assumed			-,462	290,908	,645
	MathTest	Equal variances assumed	,042	,838	-1,673	292	,095
		Equal variances not assumed			-1,674	291,664	,095
	ReadinessTest	Equal variances assumed	,072	,789	-,644	292	,520
		Equal variances not assumed			-,644	291,028	,520
Luganda	LanguageTest	Equal variances assumed	1,967	,161	1,637	562	,102
		Equal variances not assumed			1,638	560,426	,102
	MathTest	Equal variances assumed	,000	,992	-1,018	562	,309
		Equal variances not assumed			-1,018	562,000	,309
	ReadinessTest	Equal variances assumed	1,509	,220	1,130	562	,259
		Equal variances not assumed			1,131	560,440	,259
Luo	LanguageTest	Equal variances assumed	1,293	,257	-1,022	180	,308
		Equal variances not assumed			-1,028	179,520	,305
	MathTest	Equal variances assumed	3,085	,081	-1,871	180	,063
		Equal variances not assumed			-1,880	179,921	,062
	ReadinessTest	Equal variances assumed	1,288	,258	-,818	180	,414
		Equal variances not assumed			-,826	176,963	,410
Rukiga	LanguageTest	Equal variances assumed	,232	,631	1,126	258	,261
		Equal variances not assumed			1,127	257,935	,261
	MathTest	Equal variances assumed	1,333	,249	,867	258	,387
		Equal variances not assumed			,868	256,620	,386
	ReadinessTest	Equal variances assumed	1,845	,176	-,373	258	,710
		Equal variances not assumed			-,373	257,606	,709
Runyanko	LanguageTest	Equal variances assumed	,001	,979	-,132	273	,895
		Equal variances not assumed			-,132	272,582	,895
	MathTest	Equal variances assumed	,064	,800	-,909	273	,364
		Equal variances not assumed			-,910	270,954	,364
	ReadinessTest	Equal variances assumed	,007	,934	-,383	273	,702
		Equal variances not assumed			-,384	271,369	,702

Annex 13 (cont'd): T-test by sex vs other variables

Grouping variable: region

Independent Samples Test

Region			Levene's Test for Equality of Variances		t-test for Equality of Means		
			F	Sig.	t	df	Sig. (2-tailed)
Central	LanguageTest	Equal variances assumed	2,377	,124	1,403	598	,161
		Equal variances not assumed			1,403	595,780	,161
	MathTest	Equal variances assumed	,029	,864	-1,146	598	,252
		Equal variances not assumed			-1,146	597,859	,252
	ReadinessTest	Equal variances assumed	,581	,446	,614	598	,540
		Equal variances not assumed			,614	597,158	,540
East	LanguageTest	Equal variances assumed	3,242	,072	,583	577	,560
		Equal variances not assumed			,583	574,354	,560
	MathTest	Equal variances assumed	,166	,684	-,126	577	,900
		Equal variances not assumed			-,126	576,457	,900
	ReadinessTest	Equal variances assumed	,267	,605	,354	577	,723
		Equal variances not assumed			,354	574,317	,724
North	LanguageTest	Equal variances assumed	,670	,413	-1,075	586	,283
		Equal variances not assumed			-1,075	584,694	,283
	MathTest	Equal variances assumed	,004	,950	-2,980	586	,003
		Equal variances not assumed			-2,980	584,561	,003
	ReadinessTest	Equal variances assumed	2,184	,140	-1,291	586	,197
		Equal variances not assumed			-1,291	585,849	,197
West	LanguageTest	Equal variances assumed	1,833	,176	,793	555	,428
		Equal variances not assumed			,794	554,722	,428
	MathTest	Equal variances assumed	1,475	,225	-,165	555	,869
		Equal variances not assumed			-,165	554,233	,869
	ReadinessTest	Equal variances assumed	1,161	,282	-,541	555	,589
		Equal variances not assumed			-,541	555,000	,589

Annex 13 (cont'd): T-test by sex vs other variables

Grouping variable: district

Independent Samples Test

District			Levene's Test for Equality of Variances		t-test for Equality of Means		
			F	Sig.	t	df	Sig. (2-tailed)
Gulu	LanguageTest	Equal variances assumed	1,586	,209	-1,444	290	,150
		Equal variances not assumed			-1,447	286,461	,149
	MathTest	Equal variances assumed	,576	,448	-2,833	290	,005
		Equal variances not assumed			-2,843	287,777	,005
	ReadinessTest	Equal variances assumed	5,635	,018	-1,514	290	,131
		Equal variances not assumed			-1,525	289,676	,128
Kabale	LanguageTest	Equal variances assumed	,204	,652	1,164	258	,245
		Equal variances not assumed			1,164	257,735	,245
	MathTest	Equal variances assumed	1,620	,204	,900	258	,369
		Equal variances not assumed			,900	255,652	,369
	ReadinessTest	Equal variances assumed	1,995	,159	-,323	258	,747
		Equal variances not assumed			-,323	257,087	,747
Kumi	LanguageTest	Equal variances assumed	,313	,576	,182	277	,855
		Equal variances not assumed			,182	275,427	,855
	MathTest	Equal variances assumed	,063	,802	-,146	277	,884
		Equal variances not assumed			-,147	276,437	,884
	ReadinessTest	Equal variances assumed	,627	,429	-,078	277	,938
		Equal variances not assumed			-,077	269,334	,938
Lira	LanguageTest	Equal variances assumed	,159	,690	-,405	294	,686
		Equal variances not assumed			-,405	293,230	,686
	MathTest	Equal variances assumed	,321	,571	-1,516	294	,131
		Equal variances not assumed			-1,517	293,941	,130
	ReadinessTest	Equal variances assumed	,030	,864	-,612	294	,541
		Equal variances not assumed			-,612	293,163	,541
Mbarara	LanguageTest	Equal variances assumed	,017	,896	,049	295	,961
		Equal variances not assumed			,049	294,977	,961
	MathTest	Equal variances assumed	,181	,671	-1,009	295	,314
		Equal variances not assumed			-1,009	293,781	,314
	ReadinessTest	Equal variances assumed	,036	,850	-,344	295	,731
		Equal variances not assumed			-,344	294,352	,731
Mpigi	LanguageTest	Equal variances assumed	,218	,641	1,682	298	,094
		Equal variances not assumed			1,683	297,931	,093
	MathTest	Equal variances assumed	,051	,822	-,959	298	,338
		Equal variances not assumed			-,958	295,726	,339
	ReadinessTest	Equal variances assumed	,210	,647	,084	298	,933
		Equal variances not assumed			,084	295,340	,933
Mukono	LanguageTest	Equal variances assumed	3,365	,068	,330	298	,742
		Equal variances not assumed			,329	291,446	,742
	MathTest	Equal variances assumed	,300	,585	-,872	298	,384
		Equal variances not assumed			-,871	294,289	,385
	ReadinessTest	Equal variances assumed	2,280	,132	,767	298	,443
		Equal variances not assumed			,766	292,223	,444
Soroti	LanguageTest	Equal variances assumed	3,937	,048	,487	298	,627
		Equal variances not assumed			,487	297,368	,626
	MathTest	Equal variances assumed	,774	,380	-,132	298	,895
		Equal variances not assumed			-,132	297,929	,895
	ReadinessTest	Equal variances assumed	,001	,972	,346	298	,729
		Equal variances not assumed			,346	297,926	,729

Annex 14: School statistics

All numbers reported by Head Teacher during interview; 101 of 117 schools.

Region	School	Number girls	Total pupils	Number women	Total teachers
Central		no response	no response		
	Bamusuuta P/S			9	14
	Basiima P/S	332	628	9	14
	Bishop East P/S	305	573	6	14
	Buikwe Self Help Nursary P/S	134	298	4	11
	Buwoola C/U P/S	206	393	6	9
	Kawolo C/U P/S	234	485	10	13
	Kayenje CPS	245	458	5	9
	Kiryowa UMEA P/S	165	312	5	6
	Kiyoola C/U P/S	219	427	4	8
	Little Angles Nursary P/S	249	438	6	14
	Lugazi East P/S	566	1079	12	24
	Luwala P/S	293	590	5	10
	Mpongo Moslim P/S	229	420	2	8
	Ndese C/U P/S	372	734	12	18
	Nsambwe Preparatory	136	247	4	10
	Seeta Parents P/S	no response	no response	13	14
	St.Peter Nursary P/S	216	435	6	12
	Vicentalex P/S	133	285	9	16
East	Agora P/S	373	776	8	11
	Akubui P/S	338	638	4	12
	Amen P/S	148	324	5	9
	Angai P/S	336	653	2	10
	Aukot-Agule P/S	101	211	2	4
	Awoja P/S	761	1517	10	22
	BKC DEM. School-Ngora	145	314	3	8
	Go Standard P/S	198	444	11	23
	Hope P/S	126	269	11	17
	Jameler P/S	198	465	9	23
	Kabata P/S	323	651	5	10
	Kakures P/S	332	685	3	12
	Kamod P/S	477	906	5	15
	Kanyumu P/S	333	704	3	13
	Katilekori P/S	289	592	2	11
	Katine P/S	330	653	4	12
	Kings kids P/S	137	324	3	14
	Kombecce Standard P/S	80	180	2	8
	Kyere Township P/S	445	852	6	17
	Ngora Township	388	731	7	14
	Odwarat P/S	298	582	5	13
	Omerein	223	421	3	9
	Oteteen P/S	250	520	3	8
	Pioneer	524	939	18	27
	Serere T/S P/S	506	973	9	16

Region	School	Number girls	Total pupils	Number women	Total teachers
	Soroti Demonstration	670	1308	12	22
	Soroti Municipal Council	282	550	5	23
	Teso Parents	222	424	8	21
	The Light P/S	128	264	5	11
North	Abole P/S	303	592	4	10
	Bright Valley P/S	300	539	10	18
	Bungatira P/S	510	1083	12	32
	Grace Christian Academy	485	944	7	17
	Layibi P/S	471	993	14	21
	Mother Angioletta	210	474	5	15
	Negri P/S	536	536	4	16
	Omoti P/S	141	285	0	3
	Pakwelo P/S	694	1529	20	35
	Pece P.7 school	1162	2306	24	41
	Unifat	612	1217	14	34
	Wilul P/S	166	358	1	6
	Agurepenge P/S	no response	no response	1	9
	Akwete P/S	389	856	1	12
	Akwi Aworo P/S	355	700	5	10
	Awali P/S	460	663	1	12
	Fountain P/S	231	453	12	27
	Integrated P/S-Lira	281	500	9	31
	Ireda P/S	893	1727	13	29
	Lira Army P/S	559	1092	10	27
	Lira Central	838	1580	23	63
	Ocubu P/S	300	651	1	10
	Ojwina P/S	698	1320	14	30
	Omito P/S	702	1407	13	25
	St.Kizito P/S	590	1088	6	38
	St.Mary's P/S	184	424	4	15
	St.Paul P/S	512	1039	8	22
West	Allied P/S	117	322	9	22
	Bikomero P/S	98	202	2	7
	Buranga P/S	433	901	10	21
	Home Care P/S	297	630	3	19
	International Window School	307	595	6	20
	Itali Soldare Rushaki	161	299	6	17
	Jack and Jill Nursary p/s	123	225	7	13
	Junction P/S	270	550	6	12
	Kabale P/S	1107	1865	19	36
	Kabale Parents P/S	158	345	7	14
	Kabatereine Memorial P/S	360	700	8	18
	Kagoma P/S	110	248	1	10
	Kibuba	176	303	4	10
	Kigumira P/S	302	589	2	9
	Kikonkoma P/S	117	243	4	8

Region	School	Number girls	Total pupils	Number women	Total teachers
	Kyehinde P/S	327	678	2	15
	Mandera Junior School	212	334	7	19
	Mbarara Army P/S	721	1430	20	35
	Mengo P/S	275	531	2	11
	Nyakayojo 1 P/S	158	350	6	11
	Rubindi Parents	218	436	3	20
	Rukindo P/S	62	106	5	7
	Rweibogo P/S	133	262	3	7
	Rwentamu P/S	356	673	4	15
	St.Mary's P/S-Katete	202	361	4	11
	Uganda Martyrs P/S	545	1199	19	35
Total		**32922**	**64405**	**705**	**1669**	**32922**
Average		339	664	7	17	339