BMZ Federal Ministry for Economic Cooperation and Development

Future-Makers. Building the future. Let's join forces.

Learning Outcomes Assessments and Numeracy

With Reference to Early Grade Numeracy in Low Income Countries

Published by giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

BMZ Federal Ministry for Economic Cooperation and Development

Future-Makers. Building the future. Let's join forces.

Learning Outcomes Assessments and Numeracy

With Reference to Early Grade Numeracy in Low Income Countries

Published by giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Acknowledgments

The authors thank the German Federal Ministry for Economic Cooperation and Development (BMZ) and the Global Partnership for Education (GPE) for their support as well as the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH as implementing agency of BMZ.

In particular, we thank the following people:

Dr. Michael Holländer
Senior Education Advisor
Education Section
Division of Education, Health and Social Protection

Debashree Roy
Junior Education Advisor
Education Section
Division of Education, Health and Social Protection

Lena Mächel
Intern
Education Section
Division of Education, Health and Social Protection

Portions of this study were presented at the BMZ/GPE Numeracy Conference in Berlin on December 3-4, 2012 as well as during a webinar on December 12, 2012 through the Numeracy Community of Practice coordinated by the Global Partnership for Education (GPE).

This study was commissioned by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) Education Section on behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ). The views expressed are those of the author and do not necessarily reflect the position of the GIZ and BMZ.

Acronyms

ASER	Annual Status of Education Report
AusAID	Australian Agency for International Development
BMZ	German Federal Ministry for Economic Cooperation and Development
CoP	Community of Practice
DfID	Department for International Development
DIF	Differential Item Functioning
EGMA	Early Grades Mathematics Assessment
EGRA	Early Grades Reading Assessment
EU	European Union
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
GPE	Global Partnership for Education
IEA	International Association for the Evaluation of Educational Achievement
LLECE	Latin American Evaluation of the Quality of Education
LMTF	Learning Metrics Task Force
MLA	Measuring Learning Achievement
MLSC	Mathematics Learning Study Committee
NGACBP	National Governor's Association Center for Best Practices
PASEC	CONFEMEN Program on the Analysis of Education Systems
PDA	Personal Data Assistant
PIRLS	Progress in International Reading Literacy Study
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SERCE	Second Regional Comparative and Explanatory Study
TEMA	Test of Early Mathematics Ability
TIMSS	Trends in Mathematics and Science Study
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Education, Scientific, and Cultural Organization
USAID	U.S. Agency for International Development
UWEZO	"Capability" in Kiswahili

Introduction

In October 2012, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH on behalf of BMZ commissioned School-to-School International to conduct a desk study on the *Assessment of Early Grade Numeracy*, particularly in low income countries, as part of a series of studies on numeracy. Two other studies would examine *Early Grade Development and Numeracy* and *Mobile Learning and Numeracy*.

In the terms of reference, GIZ highlighted the need for these studies by stating that numeracy has been a relatively "ignored facet" of the international Education For All agenda. Numeracy has not received the attention it deserves in spite of the fact that strong competence in numeracy is necessary for full participation in society, including as the foundation for science, technology, engineering, and business. As stated by the Mathematics Learning Study Committee (2001),

> *"Citizens who cannot reason mathematically are cut off from whole realms of human endeavor. Innumeracy deprives them not only of opportunity but also of competence in everyday tasks. All children must learn to think mathematically, and they must think mathematically to learn."*

The numeracy that children develop in the primary grades forms the foundation of the knowledge and skills needed for success in secondary school and beyond. Helping students get started in developing conceptual understanding, mastery, and fluency with numeracy in the early grades is essential to ensuring that they will have the opportunity to successfully complete more advanced mathematics courses.

In many classrooms around the world, mathematics instruction is dominated by procedures, particularly in teaching students how to calculate. While calculating is a vital skill, it is often taught mechanically and without building a mastery of number concepts, i.e., what numbers represent, how they can be ordered and counted, and how they can be composed, decomposed, and compared. The emphasis on computational procedures is accompanied by neglect of other areas of numeracy, including measurement, patterns, and data and data displays. Such a narrow curriculum often limits students' ability to reason and solve problems.

Assessing children's numeracy suffers from the same issues as those found in mathematics instruction. Most assessment depends on the curriculum as a source for developing tasks. If the curriculum emphasizes calculation, for instance, then the assessments generally measure this subdomain. The linkage between curriculum, instruction, and assessment can perpetuate rote memorization of facts and procedures, without helping children learn basic mathematical concepts. The most successful educational systems are able to include broader subdomains, such as understanding patterns, interpreting data, and solving problems, and their assessment systems are able to measure them.

The assessment systems themselves are often limited to producing summative information. Traditional end-of-cycle public examinations are often the main type of assessment used in low income countries, since these assessments fill the need for selecting students to participate in the next cycle of education. However, there are other assessments that can promote system evaluation and better student learning. An increasing number of countries are organizing their own sample-based national assessments, with accompanying surveys, to provide information on whether students at particular grade levels are making progress in key subject areas. International assessments, whether worldwide or regional, are also contributing valuable information. Diagnostic assessments are also becoming increasingly popular, particularly in filling a gaping need in the early primary grades. Perhaps the area of greatest need is in the area of classroom-based assessment. Many researchers believe that this area has the greatest potential for helping teachers and students to improve learning outcomes.

The focus of this desk study is to provide information on the assessment of early numeracy learning outcomes as an integral part of efforts aimed at increasing education quality in low income countries. In order to provide focus for the study, GIZ identified three assessment-related challenges, which are crosscutting for each of the sections of the study:

1. Using assessment to support children in mastering foundational concepts and competencies, such as number sense and computation.
2. Relating or adapting assessment to the environments of children in low income countries.
3. Applying assessment to understand the strengths and weaknesses of the informal mathematics that children bring to school.

Four guiding questions, also provided by GIZ, form the main structure for the desk study. Each questions is addressed in an individual section, though there is some overlap between the questions across the sections. These questions are the following:

1. What kinds of numeracy learning outcomes need to be assessed?
2. What are the current practices in assessment of early grade numeracy learning outcomes?
3. What (low-cost, effective, and easily applicable) early grade numeracy assessment tools are available?
4. What can be learned from the assessment of early grade literacy with reference to numeracy?

Research in numeracy and mathematics education is cited throughout the desk study, and examples from low income countries are provided to illustrate many of the points. The study concludes with some final remarks on assessing numeracy in the early grades.

Table of Contents

Question 1: What kinds of early grade numeracy learning outcomes need to be assessed?

In order to decide what kinds of numeracy learning outcomes to evaluate, we have identified three different factors that influence the selection of the outcomes. These are universal vs. specific competencies, concepts vs. skills in numeracy, and informal vs. formal abilities.

We begin with a discussion of whether numeracy outcomes are universal or specific. Clearly, we need to know whether children from across the globe develop certain competencies and whether other competencies are place-specific and dependent on contexts. It is important to determine whether competencies fall into these categories when designing assessment tasks, whether for international purposes or for a particular country or region.

Similarly, we need to ensure the assessment of both concepts and skills. For instance, children need to learn the concepts behind number, place value, and operations if they are going to understand arithmetic. They also need the skills to implement those concepts, such as the procedures for computing sums, differences, products, and quotients. The assessments should be designed to find out whether children have the requisite understandings or whether their knowledge and skills are more rote and less transferable.

We also need to have a clearer understanding of children's informal and formal mathematical abilities. Informal mathematics generally pertains to those concepts and skills that children learn outside of school and in everyday life, such as basic numbers and ideas of counting. Formal mathematics involves more symbolism and abstraction, such as the algorithms children learn at school for operations. Both need to be assessed, as both are fundamental in how children learn to think mathematically.

After discussing these three factors – universal vs. specific, concepts vs. skills, and informal vs. formal – we then present different assessment frameworks currently used by various standards and assessment programs. Finally, we examine whether learning outcomes are different for students in low income countries and we offer possible explanations for the results based on the three factors.

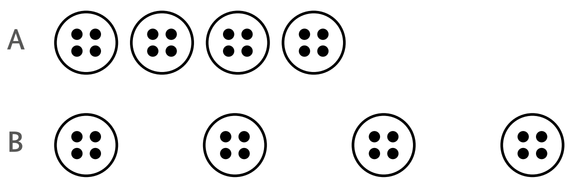
A. Universal vs. specific competencies

In order to make decisions about which numeracy competencies to assess, it is important to consider whether a competency is universal to all individuals or specific to a cultural group. Basically, properties of quantity (e.g., conservation, composition and decomposition, and order relations) are universal; how cultural groups represent these properties is specific. When designing assessment tasks, particular attention should be paid to how competencies are being represented on the test, recognizing that the representation of particular tasks may be different from what some children are taught, and thus they may fail to elicit children's knowledge of early mathematics. Below, we detail examples of universal competencies, and then discuss how systems of representation differ.

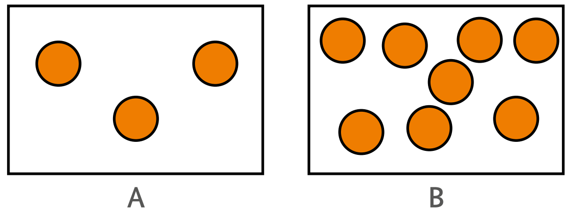
Universal

Fundamental properties of quantity are generally considered to be universal numerical competencies (Gelman & Gallistel, 1978; Ginsburg & Opper, 1988). Conservation of quantity is the idea that quantity remains constant despite changes to the size or shape of the arrangement (Piaget, 1965).

To illustrate the competency of conservation of quantity, in the figure below, both Row A and Row B have four buttons, although the arrangements are different. Across countries, young children often believe that changes to the arrangement of objects in a set change the quantity of the set. A young child may say that Row B has more because it is longer than Row A. In order to develop early numeracy competencies, however, children must learn that changes in arrangement do not change the quantity.



Another universal competency is subitizing, which is the ability to instantly recognize small quantities of number without counting them. Studies have shown that humans can subitize up to 6 objects; young children can often subitize 3 to 4 objects (Clements, 1999). In the figure below, we can subitize the dots in box A, knowing that there are 3 dots without having to count them. In box B, there are too many dots to subitize. Subitizing is an important universal competency because it allows children to be able to know the numerosity of a set without needing to know how to count. It also allows young children to compare quantities, again without needing to know how to count. Subitizing is an essential skill to measure since it is a "fundamental skill in the development of students' understanding of number" (Baroody, 1987).



Composition and decomposition of quantity is another fundamental property of number. Any quantity can be decomposed into smaller quantities, and can be combined with other quantities to make a new quantity. For example, consider a set with 5 buttons in it. The set can be decomposed into 2 and 3 buttons, or 4 and 1 buttons. The set of 5 buttons can also be combined with a new set of 3 buttons to make 8 buttons. Composition and decomposition of quantity form the basis of operations. Children learn about composition and decomposition of number as they play with objects and group and regroup them. They come to see that adding and taking objects away from sets is the only way to change the quantity of the set. Assessments can determine whether children have acquired this vital skill (Ginsburg & Baroody, 2003).

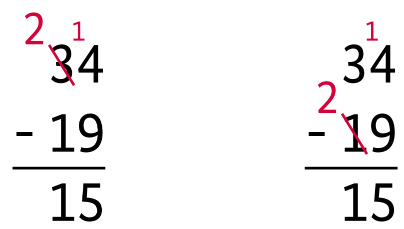
Quantity also has order relations of more than/less than embedded in it. The quantity of 5 is less than the quantity of 6, but more than the quantity of 4. Children, through subitizing, counting, and eventually number word sequences, compare quantities to determine more than and less than relationships. Two sets of dots or, for older children, a mental number line can be used to assess this skill (Ginsburg & Baroody, 2003).

Specific

Representations of quantity, as well as the different procedures and strategies we use to solve numerical problems, are often specific to local contexts. One of the primary ways that quantity is represented is through language. On a surface level, we know that different cultural groups use different words to represent different quantities. On a deeper level, some languages contain properties of a number system. For example, in the Japanese language, the numbers "11" and "12" are composed of the words for "10 and 1" and "10 and 2", respectively. In contrast, "eleven" and "twelve" are effectively nonsense words in the English language. Not surprisingly, research has shown that children in Japan learn number and place value concepts earlier than children in the U.S. (Miura et al., 1993). In an extreme case, Gordon (2004) found that members of the remote Pirahà tribe in Brazil use a "one-two-many" language system to convey quantity, and do not have other counting words; certainly, the representation of numbers beyond one and two is problematic for this population. Interpretation of assessment results should take into consideration whether the language used to represent number is easy or difficult for the children in those cultures.

There are other differences in representation of number. In some contexts, number lines have arrows on both ends; in other contexts, arrows are only on the right side. Both number lines convey they idea that numbers increase and decrease in value infinitely in different ways. Symbols for multiplication and division differ, as well as the ways people write equations (horizontal and/or vertical). Tests with items that misrepresent these symbols for particular groups of children will not lead to accurate assessments of their abilities.

Strategies to solve problems are also different. Algorithms, or mathematically proven means to solve a problem, differ by country and culture. The figure below shows the algorithm for 2-digit subtraction used in the U.S. In this problem, a "10" is borrowed from the left column (the tens) and added to the right column (the ones) to change the "4" into "14." The "3" in the tens column becomes a "2" because 10 has been borrowed. In contrast, the subtraction algorithm in Lebanon involves adding to both numbers. As in the U.S., 10 is added to the top number, converting the "4" into "14", but instead of borrowing 10 from 30, it is added to the bottom number, changing the "1" into "2". The result is the same but the algorithm varies, which might need to be taken into account when designing an assessment task.



Algorithm in the U.S. Algorithm in Lebanon

Other strategies, such as informal problem solving, may also differ (Carpenter et al., 1996). For example, when solving 29 + 34, a child may add two sets of 30 and then one set of 3 to determine the answer. Alternatively, a child may count up from 29 by tens and ones (29, 39, 49, 59, 60, 61, 62, 63). While the answers are identical, the strategies to arrive at the answer often vary according to experiences in local contexts, which may be reflected on a mathematics assessment if the task involves identifying a particular approach.

It is important to consider the ways in which competencies are represented on assessments, which should take into account what we know about cultural differences when considering the numeracy competencies of the child.

B. Concepts vs. skills in numeracy

There are concepts and skills underlying early numeracy competencies, and it is important to assess both. The Mathematics Learning Study Committee (2001) states that in order to be productive members of society, it is vital that young students understand the mathematics they are learning. Research has shown that children need to be able to make sense of mathematics in order to be successful in learning mathematics, which involves much more than memorizing a procedure or simple arriving at the correct answer. Instead, understanding mathematics is about being able to justify answers, generalize understandings to new problems, and understand why a particular answer is correct (NGACBP, 2010).

Concepts

Conceptual understanding can be thought of as making connections between pieces of knowledge. According to Piaget (1964), there are two processes involved in making these connections. One is assimilation, which involves connecting new information to existing knowledge. For instance, a child who has knowledge of addition of objects can learn the "+" symbol so that the operation can become symbolic. The other is integration, which involves connecting two existing pieces of knowledge. For instance, a child who can recognize that there are five fingers on each hand and ten fingers on both hands will at some point realize that the sum of five and five is ten. Assessing whether a child has made these kinds of connections will tell us whether the concept has been acquired, and also whether the child is ready to learn more advanced concepts.

Another way to look at numeracy concepts is that they involve knowing "why" or the reason(s) behind certain mathematical knowledge. As stated in the Common Core State Standards (NGACBP, 2010):

> *"One hallmark of mathematical understandings is the ability to justify, in a way appropriate to the student's mathematical maturity, why a particular mathematical statement is true or where a mathematical rule comes from."*

For example, a child solving a simple addition problem of 3 + 4 needs to know more than just the answer to the problem; instead, the child should also be able to prove that their answer is correct, either by demonstrating the strategy they employed to solve the problem or by reasoning about the meaning of addition. Hatano (1988) uses the term "routine expertise" to describe the rote knowledge of a task, which is much less powerful than the "adaptive expertise" that allows the child to apply the insight to both familiar and new problems. Well-designed assessment tasks can tell us whether a child, for instance, has memorized a number fact or whether the reasoning behind the number fact has been grasped.

Skills

Ginsburg et al. (1998) define numeracy skills as "how to" solve a problem. Skills are the procedures, number facts, and formulas for solving problems. In the example above of 3 + 4, a child may have memorized the number fact and know the answer is 7. Alternatively, the child may count up using their fingers.

Skill fluency requires being able to carry out computations efficiently, appropriately, and effectively (MLSC, 2001). It also means knowing when to use the skills and how to adapt them to new circumstances. Knowing "how to" and "when" are important components of learning mathematics and should be part of the tasks in any assessment of early numeracy.

Concepts and skills are closely tied to each other, and therefore both are needed in assessments (Wu, 1999). In order to develop early numeracy competencies, a child has to know both how to solve a problem as well as why the answer is correct. Assessments that focus only on skills may leave invisible children's understanding of numeracy concepts, and those that concentrate only on concepts will not provide information on whether a child can apply mathematical understanding. Assessments should focus on both concepts and skills in order to provide the most accurate information possible about a child's competencies.

C. Informal vs. formal abilities

Both informal and formal concepts and skills are fundamental for children in learning to think mathematically. As children interact with their world, they learn informal mathematical skills, such as counting, grouping and regrouping, and using operations to solve simple problems. In school, children are introduced to formal mathematics, where quantities are represented by numerals, and mathematical thinking is largely abstract and symbolic. The informal knowledge the children acquire primarily outside school forms a strong base for learning formal mathematics.

Informal

Informal numerical competencies are those learned by children in their everyday lives through interactions with adults, peers, and objects in the world (Piaget, 1965; Nunes, Schliemann & Carraher, 1993; Sitabkhan, 2009). Children develop these competencies prior to formal schooling because "it is personally meaningful, interesting, and useful to them" (Ginsburg & Baroody, 2003). Informal numeracy competencies are:

1. Built on counting and invented strategies to solve problems.
2. Embedded in practical activities such as sports or shopping.
3. Reflective of mathematics in a local context.

Informal strategies to solve problems are based largely on counting and invented strategies. These strategies are often oral instead of written. In their everyday lives, children engage in multiple activities that involve numeracy. For example, children may count pieces of fruit in a market or the number of steps they take to walk from their home to their school. They may solve simple problems, such as figuring out how many cups are needed for each member of their family. Children invent procedures to solve these kinds of problems, largely based on counting.

Informal numeracy concepts are concrete and meaningful and embedded in children's activities (Guberman, 1996; Nasir, 2001; Sitabkhan, 2009). Children build understandings that have real-world analogies. For example, children may need to calculate the total cost of items they are buying in a store (Taylor, 2009), or calculate the change given to customers in a market setting (Sitabkhan, 2009). The mathematical problems they solve refer to real situations and objects in the world.

Because informal concepts are embedded in the everyday world of the child, the concepts often reflect understandings of the local context. For example, Baranes et al. (1989) found that children in the U.S. were more successful solving addition and subtraction problems with the number "25" than children in other countries since the quarter (25 cents) is a common coin. Saxe (1991) studied young candy sellers' mathematical understandings in Brazil and found that they were able to recognize numerals on bills and coins but not when the same numerals were written on paper.

It is critical to assess informal mathematics because school children assimilate or integrate what they are taught in school largely in terms of the mathematics that they learn outside of school. If the assessment tasks are not well designed, it is likely that strong informal knowledge and skills will not be discovered. In particular, teachers must use classroom assessment to form an accurate picture of children's informal abilities so that instructional strategies can help children make connections with formal mathematics. If gaps in informal knowledge are not remedied, children will likely have difficulties with the corresponding formal mathematics topics. On the other hand, assessments of informal mathematics can also help prevent the teacher from wasting valuable time by avoiding instruction on competencies that the child has already mastered (Ginsburg & Baroody, 2003).

Formal

Formal competencies are those that are taught in formal schooling. They are usually abstract and based on interpretation of symbols and mathematical representation. Formal competencies in early mathematics include the ability to recognize numerals, demonstrate knowledge of number facts, identify simple patterns, and solve written addition and subtraction problems. Early competencies form the basis for more advanced concepts and skills, such as multiplication, division, fractions, complex patterns, and data interpretation. Formal competencies are usually written, not oral, and involve understanding and making sense of symbols and representations.

Learning facts, procedures, and formulas must be tied to conceptual understanding, much of which is informal. Attaching meaning to learning makes tasks such as memorizing facts and definitions easier since the child can make sense of the abstract or symbolic representations. Children who forget number facts, for instance, can reconstruct them by using their informal mathematics and conceptual understandings. This also applies in situations where the child encounters an unfamiliar task. They are less intimidated by new problems if they have the background to solve them (Ginsburg & Baroody, 2003).

Assessing formal mathematics should be accompanied, when possible, by tasks that assess the foundational knowledge and skills from informal mathematics. It is even possible to assess the same numeracy competency both formally and informally. For example, in assessing addition, a child can be given a set of 4 and a set of 5 concrete objects and asked to determine how many there are all together (informal); the child can also be given the written problem of 4 + 5 and asked to solve it (formal).

By focusing on both formal and informal, we can make better decisions about children's abilities by finding out about their strengths as well as diagnosing their difficulties. Taking the addition example from above, administering the informal task can reveal that a child understands and can correctly solve simple addition problems, but the tasks focusing on formal competencies can reveal that the same child cannot correctly solve a written addition problem. A conclusion from this might be to focus instruction on abstract numeracy understandings, helping the child understand the symbols. In this way, instruction can also use informal knowledge to build formal knowledge.

D. Frameworks, standards, and tasks for assessing numeracy

We now turn to the different competencies that are identified in well-known frameworks and standards. These include TIMSS, Common Core, and UNESCO/UIS/Brookings (LMTF – Learning Metrics Task Force). Also listed are the outcomes measured by tasks on some of the most popular international assessments of early numeracy, such as EGMA, TEMA, ASER, and UWEZO. To varying degrees, the frameworks, standards, and tasks measure competencies that are universal and specific, conceptual and skills-based, and informal and formal.

1. Trends in International Mathematics and Science Study (TIMSS) Framework
   - Number
   - Geometric Shapes and Measures
   - Data Display

2. Kindergarten-Grade 2 Common Core State Standards in the U.S.
   - Counting and Cardinality
   - Operations and Algebraic Thinking
   - Number and Operations in Base-10

3. UNESCO/Brookings International Primary School Standards
   - Number and Operations
   - Mathematics Applications
   - Geometry and Patterns

4. Early Grades Mathematics Assessment (EGMA) Tasks
   - Number Identification
   - Quantity Discrimination
   - Patterns and Missing Numbers
   - Addition and Subtraction (Facts and Algorithms)
   - Addition and Subtraction (Word Problems)
   - Relational Reasoning
   - Spatial Reasoning

5. Test of Early Mathematics Ability (TEMA) Tasks
   - Numbering (Informal)
   - Number Comparisons (Informal)
   - Calculation (Informal and Formal)
   - Concepts (Informal and Formal)
   - Numeral Literacy (Formal)
   - Number Facts (Formal)

6. Annual Status of Education Report (ASER) Tasks
   - Number Recognition
   - Addition and Subtraction (Facts and Algorithms)
   - Addition and Subtraction (Word Problems)
   - Multiplication and Division (Facts and Algorithms)

7. "Capability" in Kiswahili (UWEZO) Tasks
   - Number Recognition
   - Addition and Subtraction (Facts and Algorithms)
   - Addition and Subtraction (Currency)
   - Multiplication and Division (Facts and Algorithms)

A comparison of the frameworks, standards, and tasks shows that numbers (i.e., counting, numerals, comparisons, place value, facts, quantities, and patterns) and operations (i.e., addition, subtraction, multiplication, and division) are found in most of the lists. Areas such as geometry, spatial reasoning, and data, are less well represented. On most of the assessments, the vast majority of tasks are universal, skills are emphasized over concepts, and formal mathematics is predominant over informal mathematics.

E. Numeracy outcomes for children in less developed countries

Children from low income countries may not perform as well on assessments of formal concepts/procedures as measured by international assessments. Results from TIMSS show that children in low income countries generally score below the mean in primary school mathematics (Mullis et al., 2012). This was particularly the case in mathematics tasks that assessed the cognitive areas of applying and reasoning. Accompanying surveys of students showed that performance was related to the following factors, all of which were at lower levels in socio-economically less developed countries:

- Early numeracy activities before the beginning of primary school;
- Attendance in pre-primary education (preschool) programs;
- Ability to do numeracy-related tasks upon entry into primary school;
- Home resources (e.g., parental education level, books) for learning;
- Ability to speak the language of assessment as a first language;
- Mathematics resources (specialized teachers, materials) at the school;
- School discipline, safety, and orderly environments;
- Teacher preparation and confidence in teaching mathematics; and,
- Teacher experience level and career satisfaction.

Other international studies, including those that have featured one-on-one interviews and oral assessments, have shown that children in low income countries show strengths in the kinds of numeracy knowledge and skills – i.e., specific and informal skills – that are not generally measured on international assessments (Ginsburg & Baroody, 2003). Children either in low income countries or in high poverty areas of middle income countries have demonstrated strengths in areas such as comparing quantities, mental arithmetic, and operations with currency, particularly when presented with problems from their everyday lives (Carraher, Carraher & Schliemann, 1985; Davis, 1992; Saxe & Posner, 1983).

Another issue affecting assessment results is that children in low income countries may learn at rates that are different from those of children in high income countries. As mentioned above, issues such as learning in multiple languages can influence acquisition of some knowledge and skills. In one research study, young children in West Africa could often count to ten in their local language and in an international language, but could not count to 20 in either language (Davis, 1992). In a related study, U.S. children demonstrated higher informal skills than Asian children at younger ages, but the Asian children surpassed their American counterparts once the children reached school and began to learn formal mathematics. Both groups were compared to West African children, who performed well in both informal and formal mathematics, particularly once they reached primary school and in urban areas (Davis & Ginsburg, 1993).

The school age of the children may also influence learning. Compared to their counterparts in Western countries, children in many low income countries may be older at the same grade level. An older child might have better informal mathematics (e.g., the Piagetian skill of conservation), but they also might be subject to other factors, such as having to repeat classes due to poor instruction and suffering from a stigma for being older, that would negatively influence their learning (Anderson et al., 2001; Grogan, 2008).

Finally, attitudes towards mathematics can have a strong correlation with achievement. The relationship is bidirectional, with attitudes and achievement mutually influencing each other (Mullis et al., 2012). Related factors are student confidence with mathematics and student value for mathematics, both of which are highly correlated with achievement, even at grade four (Mullis et al., 2012). As mentioned above, teacher confidence with mathematics is also highly correlated with student performance. These anxieties may play themselves out in the assessments, as teachers may be afraid to assess students as they are afraid of the outcomes, and students may not want to take the assessment for fear of results. Similarly, district and country officials may not want to assess their students out of fear of the possible results.

Research has shown that student's attitudes and beliefs about themselves as math learners are related to their achievement (Stevenson, Lee & Stigler, 1986; Stigler & Fernandez, 1995). In related research, Dweck (2006) identified two kinds of mindsets: fixed and growth. A fixed mindset is one in which you are either good at something or not good at something based on your brain and genetics. A growth mindset is one in which you can become good at something based on hard work and dedication. If teachers and students have fixed mindsets, this may lead to a lack of learning in mathematics. For example, if a female student believes that only boys are good at math, she may not try as hard to learn the content during the class. When assessing mathematics, it is important to also assess student beliefs and attitudes towards mathematics and their relationship to achievement.

Summary

This overview of different factors to consider comes from studies in both mathematics education and developmental psychology. The ideas of universal vs. specific, concepts vs. skills, and informal vs. formal can have overlaps, but the general principles help us with constructing the tests, interpreting the results, and using the data to improve instruction and make policy decisions. Each situation has its own specificities that should be taken into account in assessment activities.

Particular areas of attention include the recognition that some competencies are specific to certain countries and cultures, concepts are foundational for proficiency in numeracy, and informal concepts and skills are needed to make sense of formal mathematics. These findings have applications in both assessment and instruction, particularly in low income countries where scarce educational resources need to be maximized and children's strengths in numeracy are often overlooked.

Question 2: What are the current practices in the assessment of early grade numeracy learning outcomes?

In this section, we address current practices in the assessment of early grade numeracy. First, we provide an overview of five different types of assessments used with primary school students. We broadly describe each type, including their purposes, uses, subject areas, strengths, and weaknesses. We make reference to their applicability in low income countries.

Second, we discuss each of these types of assessments in relation to program evaluation, which is one of the main uses of assessment in international educational development. In particular, the assessments are examined in terms of their suitability in measuring changes in student learning as a result of an intervention. Third, in a related topic, we discuss which of these assessments can be best used in policy dialogue to promote numeracy in low income countries.

A. Types of assessments, with their purposes, strengths, and weaknesses

Kellaghan and Greaney (2003, 2004) classified assessments in four categories: public examinations, national assessments, international assessments, and classroom assessments (also called school-based, formative, or continuous assessments). They saw assessments as satisfying four needs:

1. For describing the knowledge and skills that constitute a quality education;
2. For measuring student achievement;
3. For evaluating progress toward the goal of a quality education; and,
4. For translating assessment data into policy and instructional procedures that will improve the quality of learning.

With the recent development of short assessments for the early grades, we have added the additional category of diagnostic assessments. Each of the five categories is described below on the basis of the purpose of the assessment and of the administrative procedures for the assessment. In addition, the use of the assessment information is different for each category, both in terms of making management decisions and using the information to improve classroom instruction and student achievement.

For a donor agency, the main purpose of an assessment has traditionally been in the areas of evaluating the impacts of programs and monitoring system performance. However, the use of assessment to improve instruction and achievement has taken on increasing importance as more information has been published about the low learning levels of children in low income countries, particularly in the basic skills of literacy and numeracy. For instance, there has been some discussion of whether high-stakes examinations are promoting learning and whether enough attention is being given to classroom-based assessment, even though promoting classroom assessment necessarily implies expensive in-service teacher training and follow-up supervision and mentoring.

These issues will be discussed in the following sections. We begin by listing the five categories of assessment and providing information on their purpose/use, strengths, and weaknesses. We also cite a few examples from low income countries where appropriate.

Public examinations

1. Purpose/use: Curriculum-based, census-based, summative assessment for selection, promotion, certification, and/or retention of students; can be all subjects in the curriculum.
2. Strengths: Government certified, publically accepted, sustainable, system accountability, promotes curriculum mastery, may cover many subjects.
3. Weaknesses: Not usually focused on early grades, expensive, time-consuming, takes away instructional time, stressful, may be some issues with test security and fairness, lack technical properties to track results over time, may cover too many subjects.

4. Examples: Most low income countries, particularly where there are limited places at the upper levels of education (e.g., lower secondary, higher secondary, and tertiary).

National assessments

1. Purpose: Curriculum-based, sample-based, summative assessment for system evaluation, policy analysis, trend analysis, and instructional improvement; usually focused on language, mathematics, and science.
2. Strengths: Relatively high quality, sometimes administered at both lower and upper primary grades, often well-funded, international support, may be sustainable if funding is stable.
3. Weaknesses: Not usually focused on early grades, only administered every few years, may be some issues with capacity necessary to implement assessments without international technical support, may not be sustainable if funding is unstable.
4. Examples: Egypt, Ghana, Namibia, Pakistan, Vietnam, and Zambia.

International assessments

1. Purpose: Based on international standards, sample-based, summative assessment for system evaluation, policy analysis, international comparisons, trend analysis, and instructional improvement; usually focused on language, mathematics, and science.
2. Strengths: High quality tests, international comparisons, strong sampling, international technical and financial support (more so for the international assessments rather than the regional assessments), trend analysis, may be sustainable if funding is secure and stakeholders provide support.
3. Weaknesses: Often targeted to Grade 4 and above, expensive, may or may not be aligned with national curricula, can be demoralizing since low income countries tend to have low scores, participation may not be sustainable if public rejects findings and/or funding is unstable.
4. Examples: TIMSS (international), PIRLS (international), SAQMEC (Anglophone and Lusophone Africa), PASEC (Francophone Africa), MLA (international), and LLECE/SERCE (Hispanophone and Lusophone Latin America).

Classroom-based assessments

1. Purpose: Curriculum-based, census-based, formative assessment for examining student strengths and weaknesses, targets classroom instruction, materials development, and improved student learning; can be developed for all subjects in the curriculum, but might be better suited for a few key subjects (language, mathematics, science).
2. Strengths: Most beneficial for instructional and learning improvement, may be used for records and grades, different models depending on context, curriculum-based, should be aligned with materials development and training programs, should be sustainable if the design includes extensive training and planning for scale-up.
3. Weaknesses: Usually not standardized, quality may vary, requires teacher training, may require substantial materials development, may not be sustainable if it is not well-integrated into existing policies, programs, and systems.
4. Examples: Honduras, Namibia, and Zambia.

Diagnostic assessments

1. Purposes: Non-curriculum-based, sample- or census-based, obtain a standardized score, once or twice per year (e.g., beginning and end) for examining student strengths and weaknesses, targeting instruction, and placing students in remedial programs; focuses on a few subjects (language, mathematics).
2. Strengths: Gather data on early learning, may provide results that are parent-friendly, may impact changes in curriculum and instruction; can focus on a few indicators and/or calculate a total score with performance categories for students.
3. Weaknesses: Traditionally not part of the system in low income countries, quality of administration may vary (even if standardized), tends to be top-down.
4. Examples: EGMA (DR Congo, East Timor, Iraq, Jordan, Kenya, Liberia, Mali, Morocco, Rwanda, and Zambia), TEMA (Benin, Haiti, and Macedonia), ASER (India and Pakistan), and UWEZO (Kenya, Tanzania, and Uganda).

B. Types of assessments in relation to program monitoring and evaluation

Many of the five categories of assessments may be used for program evaluation, though they must be targeted for this purpose. More commonly, new assessment instruments and procedures are developed for program evaluation purposes since the interventions may be somewhat specialized, both in terms of content and geographical coverage. In brief, the appropriateness of the five types of assessment for program evaluation purposes is summarized below.

Public examinations

These assessments are usually not well suited for program evaluation since they are generally norm-referenced and not statistically equated across administrations. With norm-referenced assessments, percentile ranks mean that scores for one group must go down for another group to go up, so improvement is based on standing relative to other schools or districts (rather than based on a criterion, meaning that any schools can go up or down depending strictly on performance in relation to standards rather than depending on the performance of other schools).

In addition, part of the purpose of public examinations is selection to the next level of education, which implies that there are limited spaces. It would be next to impossible to show improvements in the number of passing students since that number is more or less fixed in relation to the available places at the next level of education.

Without statistical equating, scores cannot be accurately compared across administrations (since there is no equivalence in terms of difficulty of test forms).

National assessments

These assessments may be appropriate for program evaluation, since they are criterion-referenced and they are usually (or should be) equated for difficulty over time. However, the grade level(s) and subject area(s) of the national assessment must correspond with those of the program interventions, which is often not the case. Another problem is the timing of the assessments, so it is unlikely that a baseline, midline, and endline for an intervention will correspond with the dates for the national assessment. A final problem is that the sampling for the national assessment needs to correspond with the schools or districts in the intervention, which is rarely the case. In other words, the national assessments are usually on a timeline, such as every three years, while projects can start and end at any time.

International assessments

These assessments may be applicable to program evaluation, though they have the same problems as national assessments, and are even less flexible since the assessment period needs to correspond with the plans of the international planning group and the assessments of other countries. Also, again even more so than the national assessments, the international assessments have fixed subject areas, which might not correspond with the subject areas in the intervention. Another problem is that they are based on international learning standards, so the material taught in a particular country may or may not correspond to those standards. Finally, as with national assessments, the sampling needs to cover the intervention schools and districts, and the sampling plan may or may not correspond with the guidelines of the particular international assessment.

Classroom assessments

These assessments are generally not suited for program evaluation since they do not usually involve conducting standardized tests. Exceptions might be if the classroom assessment program specifies particular tests for administration at a few points in the year, such as at the end of each term. There would still be a potential problem of standardization – printing, test administration, test security, scoring, etc. Particularly well-suited would be the more recent Teacher EGRAs and EGMAs, which are designed to be administered, scored, and recorded by the classroom teacher according to standardized guidelines. They also have the potential to be aligned with the curriculum and/or project-supported interventions. A key would be to evaluate the assessments to see whether they have adequate psychometric properties. In addition, classroom assessments might have some contribution to make in qualitative program evaluations through observations, questioning, and interviews; these instruments may be combined with a Teacher EGRA in order to have both quantitative and qualitative information for a program evaluation.

Diagnostic assessments

These assessments can be used for program evaluation. Steps should be taken in the construction or adaptation of the tests so that they align with the interventions, i.e., the grade levels and specific objectives in the curriculum. They also need to be administered in a standardized format so that comparisons can be made across schools and students. The versions of the diagnostic assessments should have adequate psychometric properties (e.g., high internal consistency reliability) and content coverage (and alignment to the training, curriculum, textbooks, etc.) so that they are valid for program evaluation purposes.

C. Types of assessments in relation to policy dialogue

The five types of assessments should all be a part of the education policy of a country. The assessments must be implemented as planned so that they can fulfill their purpose. The details on each assessment and how it can be used to improve the quality of students' learning is a critical topic for discussion at the highest levels of the Ministry of Education.

Some of the current trends on the use of each type of assessment and implications for policy dialogue are discussed below.

Public examinations

These assessments have been the most important type of assessment in many countries over the past several decades. They play an important role in certifying student achievement and selecting students for higher levels of education. At the primary school level, the examinations always include a test on mathematics. In terms of early mathematics, the examinations are not applicable since they are administered at the end of the primary school cycle. The knowledge and skills that children acquire at the early grades clearly has an effect on performance at the end of primary school (Duncan et al., 2007) but the examinations do not provide a performance measure in early numeracy.

These examinations are often criticized for only measuring the kinds of knowledge and skills that students need for the next cycle, rather than covering the kinds of competencies that students need to know if they are terminating their education, i.e., not going to the next cycle, which in many countries is most of the students. For instance, there is often little coverage of everyday mathematics that children need for their life outside of school. The tests also tend to measure much rote memory and lower level cognitive skills.

The high-stakes element of public examinations can have positive and negative influences on learning. Most students study hard for the tests, but at the same time this effort spent on preparation leads to a focus on extrinsic rather than intrinsic rewards (Kellaghan & Greaney, 2003). Several months in the terminal grade levels (e.g., grade 6 for primary school) can be spent on exams preparation. Teachers often concentrate on those students who have the best chance for success rather than providing instruction for all students.

Policy makers have decisions to make in terms of the emphasis placed on public examinations. In even a medium-sized country, the examinations can cost millions of dollars. There is a necessity in most low income countries to have some kind of selection mechanism since the number of places at the next cycle of education is usually limited, with as few as one-fourth of students proceeding to secondary school in some countries. Other types of assessment are likely better in terms of promoting learning for all students, especially those who stop their schooling in primary school.

Technically, the public examinations could often benefit from better alignment with the curriculum, pilot testing so that the best items can be selected for the operational (final) test forms, test equating so that results can be compared from year to year, and better security measures. In mathematics, there should be greater links to practical numeracy and the kinds of knowledge and skills that children need to succeed outside of school and in the workplace.

National assessments

These assessments are relatively new for most countries. The focus is on the system rather than on individual students. National assessments came to the forefront in the 1990s as the result of an almost exclusive focus on educational inputs – numbers of teachers trained, numbers of textbooks produced, etc. – without information on outcomes, including the achievement levels of students and the number of students finishing a cycle of education. The implementation of national assessments was designed to provide more information to planners and policy makers on outcomes.

National assessments are often administered at an upper primary and a lower secondary grade level. For most countries, the assessments cover grades 4 or 5 and grades 7 or 8. Students are sampled at those grade levels to participate in the assessment, and the results are generalized to the population. They almost always include measures of mathematics. Data are correlated to demographics and other survey information to add a contextual element to the test scores. Unlike public examinations, national assessments offer the opportunity to look accurately at trends or changes in scores over time. This kind of information is vital to policy makers, which, in fact, is the main purpose for the assessments.

There has been a trend to use diagnostic assessments at the earlier grades, sometimes on a national basis. For most countries, the target group is grades 1 to 4. This can provide a valuable set of information on early learning, particularly in terms of identifying areas of difficulty so that interventions can be developed and implemented. This avoids past practices of waiting until the end of the cycle for large-scale assessment information, at a point in time when unfortunately it is too late to help most students.

International assessments

These assessments differ from national assessments in that they offer comparative information on a country's performance relative to that of other countries. The best international assessments also offer rich information on the curricula of different countries, such as whether a particular country is teaching a topic at the same time as in high-achieving countries. This is important in numeracy, in which children need to be challenged at particular grade levels with relevant and appropriate content.

The comparative aspect, both in terms of test scores and curricula, provide policy makers with vital information for analysis and dialogue. For instance, if a country places below its neighbors, the country has that information as a starting point, but it also knows something about the curriculum areas in which students do not perform well. Policy makers can ask for changes to the curriculum and instruction of a country. For instance, low scores in mathematics might be due to an issue with the curriculum or something that needs to be changed in instruction.

The issues cited above with national assessments generally apply to international assessments. The IEA (International Association for the Evaluation of Educational Achievement) has recently announced the development of TIMSS Numeracy 2015, which will target knowledge and skills that children should develop in lower primary school. The test would also be a paper-and-pencil complement to the orally administered diagnostic assessments that are administered in many countries. TIMSS Numeracy will provide valuable information to countries, particularly the low income countries that do not traditionally have large-scale assessments in the early grades. The IEA will produce reports that analyze policy-related factors associated with assessment results, such as curriculum coverage, teacher effectiveness, student attitudes, and parental involvement.

Classroom assessments

These assessments are not commonly used in policy dialogue since they are school-based and the information is not (and cannot be) aggregated at a national level. However, they deserve much more attention than they are currently receiving. In fact, some experts believe that classroom- or school-based assessments should take the place of public examinations at the primary school level. The negative effect of examinations could be eliminated if competencies in the curriculum can be validly assessed at the school level (Kellaghan & Greaney, 2003).

Early numeracy is one of the main subjects for classroom assessment. While most of the assessments are through daily interaction with children – observations, questioning, classwork, and homework – there are ample opportunities to administer quizzes at the end of each month. This information can be used in combination with end-of-term and end-of-year tests, to provide valid and reliable information for grade cards. Even though this process is not standardized across schools, the information can provide a valuable component for policy dialogue. An increasing number of countries are promoting school report cards as a way of informing school officials and parents of the situation in a school in a given year, which should then assist in the planning process for the following year and beyond.

Diagnostic assessments

These assessments are the most recent part of assessment programs in many countries. In spite of the relative newness of diagnostic assessments, they are rapidly being used in policy dialogue. Many of these assessments are sample-based and take 10-15 minutes to administer in a standardized format, thus the information can be aggregated to a regional and/or national level. This makes it possible to conduct policy analysis and dialogue by all stakeholders, and for policy makers to take note when numeracy levels are low.

Diagnostic assessments, along with classroom assessments, are perhaps the least well known of the five types of assessment, but they offer much potential for educational policy. Both assessments can be instrumental in gathering information at the early grades, which is the most important time to recognize children's strengths and weaknesses in numeracy and develop interventions that can provide benefits in the short and long terms.

Summary

All of the different types have some use in policy dialogue, though some more than others, and the purposes and kinds of information are different. In addition to collecting valid and reliable information, perhaps the most important message for using assessments in policy dialogue is that the information has to be timely and presented in an understandable manner. If not, policies involving assessment will have very little impact on learning. Assessment information can improve policy by providing indicators on the outcomes of education, which can reshape critical pieces of an instructional program such as the curriculum and teacher training.

Two main efforts are required for effective policy analysis and dialogue. First, the assessment systems must be coordinated at a national level and aligned with all aspects of the system – teacher training, curricula, textbooks, and the capacity to absorb students at higher levels. Second, assessment information must be processed and presented in a way that can be useful for school personnel, who then must use the information to improve instruction and learning, as well as increase parental involvement in their children's early education.

Question 3: What (low-cost, effective, and easily applicable) early grade numeracy assessment tools are available?

In this section, we select three well-known and widely used assessments (EGMA, TEMA, and ASER/UWEZO) that focus on young children's early numeracy. We provide a discussion of the needs addressed by each test, the test development process, the competencies that the instrument addresses, the types of items, administration procedures, scoring, score interpretation, strengths, weaknesses, and recommendations for improvement. We also build on the previous section of this study by listing where these assessments have been used in low income countries.

We then compare the three assessments, using a common framework as a reference point, by content area. We consider the question of whether these assessment tools are sufficient for use in developing country contexts. We discuss whether new tools are needed, or whether these tools are adaptable enough for different contexts. We provide additional discussion on the importance and need for the development and adaptation of tools for classroom-based assessment. Finally, we conclude this section with comments about technology and numeracy assessment.

A. Characteristics of low-cost, effective, easily applicable assessment tools

The three tests are presented below, with a description of the tests using the features listed in the introduction to this section.

EGMA

1. Needs Addressed: EGMA was developed by a team of consultants under the leadership of RTI International in 2010 with funding from USAID and the World Bank to address the need for an early mathematics diagnostic assessment in low income countries. It was a follow up to the success of Early Grades Reading Assessment, which was developed starting in 2005. The main purpose was to provide information that would spur policy dialogue around numeracy, and eventually interventions to improve young children's mathematics.
2. Test Development: Steps in the EGMA development process included the following: 1) an extensive literature review that established an assessment framework; 2) the formation of a group of early mathematics experts to discuss skills and tasks; 3) the development of a draft instrument, followed by piloting and revisions; 4) the implementation of the instrument (thus far in 11 countries), including adaptations for different contexts; and 5) the re-convening of the expert panel to review the implementation and make recommendations for refining the tasks and instructions.
3. Competencies and Items: Competencies measured by the instrument include both conceptual understanding and skills (i.e., procedural fluency/automaticity). There is less focus on informal mathematics. As mentioned in the sections above, the core EGMA instrument currently has seven subtasks, each of which has multiple parts. Two of the sections are timed, e.g., the child has one minute to complete the task. Note that other items involving oral counting, one-to-one correspondence, and number lines have been used in the past but are not included in the most recent core version of the test.
   - Number Identification (Timed): Reading numbers of 1- to 3-digits.
   - Quantity Discrimination: Stating which number is greater in value.
   - Missing Numbers: Filling in missing numbers by using patterns.
   - Addition and Subtraction Problems (Timed): Performing addition and subtraction calculations for both 1- and 2-digit numbers.
   - Addition and Subtraction Word Problems: Solving word problems involving all of the basic operations from real life situations.
   - Relational Reasoning: Using logic, such as the commutative property, to solve addition and subtraction problems.
   - Spatial Reasoning: Visualizing spatial patterns and mentally manipulating them.

4. Administration Procedures: The instructions for EGRA begin with telling the child about the assessment and trying to make the child feel at ease. The assessor then fills in information about the child on the instrument and begins with the items. Each item has its own set of instructions. The assessor stops the administration of the item if the child answers a successive number of parts incorrectly or if the child delays for more than a few seconds. If the item is timed, the assessor needs to use a stopwatch. On some of the items, there is a practice item so that the child has a better idea of what they are supposed to do.
5. Scoring: Each item is scored using a standard protocol. The assessor notes the item scores on a score sheet.
6. Reporting: Scores are reported in terms of number correct and number correct out of one minute (for the timed items). Some analysts have created percentage correct scores and grand means for a total score on the assessment. All scores are criterion-referenced.
7. Strengths: Strengths of the instrument include high face validity, with universal tasks that represent a progression of foundational skills; the possibility for context-specific adaptations; measurement of both concepts and skills; better reliability through oral, one-to-one administration; and proven ability to "measure the pulse" of the general level of children in early mathematics in low income countries.
8. Weaknesses: Weaknesses of the test include relatively high costs for administration, since only about 20 children can be assessed in one day by a test administrator; improbability of being able to collect data on each student, which would be useful for diagnostic assessment; lack of development on a country-specific basis, so alignment to the curriculum for individual countries must be done on a post hoc basis; lack of familiarity by most countries with oral administration, so initial discussions with education officials and training needs can be extensive; and issues with sustainability (in light of the other weaknesses).
9. Suggestions for Improvement: Suggestions for improvement are in the areas of implementation and sustainability. For instance, up to this point, there have not been examples of low income countries that have implemented the assessment completely with their own funds and personnel. There are issues with disseminating the data and using it for instructional improvement.
10. Countries of Implementation: Of the tools described in this section, the EGMA instrument has been used in perhaps the most low income countries – eleven at last check. These include DR Congo, East Timor, Iraq, Jordan, Kenya, Liberia, Mali, Morocco, Rwanda, and Zambia. Different versions of the basic seven-component EGMA have been implemented in these countries. However, validity and reliability studies are not as extensive as with TEMA and its contribution to policy dialogue is only at the beginning stages. EGMA has the same kind of potential as its sister instrument (EGRA) in educational policy analysis.

TEMA

1. Needs Addressed: The first version of the TEMA (Ginsburg & Baroody, 1983) was developed in response to two needs in early mathematics. These were for a test that would: 1) identify children in grades K to 3 with difficulties in learning mathematics; and 2) provide useful information about the mathematical strengths and weaknesses of children, both with and without learning difficulties. Since the initial publishing of the TEMA, three additional purposes for the instrument have evolved: 1) suggest instructional practices for children; 2) document children's progress in learning mathematics; and 3) serve as a measure in research projects.
2. Test Development: Steps in the development process included: 1) identify items from research studies by the authors and other researchers on children's informal and formal mathematical knowledge; 2) pilot the items; 3) develop an initial version of the instrument; 4) finalize the instrument; 5) write instructions for test administration; 6) collect reliability and validity data; and 7) create a norm group based on a representative sample of children in the U.S. In response to critiques by test reviewers, including in widely accepted publications such as the Mental Measurements Yearbook, the instrument has since gone through two major revisions. The current instrument is the TEMA-3.
3. Competencies and Items: The latest version of this instrument, the TEMA-3, has 72 items. This is an increase from the original TEMA, which had 50 items. The items measure children's informal and formal mathematics. Each item is administered individually to the children. Some of the items have multiple parts. The content domains tested by the TEMA-3 are the following:
   - Numbering (Informal): 23 items
   - Number Comparisons (Informal): 6 items
   - Calculation (Informal): 7 items
   - Concepts (Informal): 4 items
   - Numeral Literacy (Formal): 8 items
   - Number Facts (Formal): 9 items
   - Calculation (Formal): 10 items
   - Concepts (Formal): 5 items

   There are a total of 40 informal items and 32 formal items. The TEMA documentation gives descriptions of each of the item domains and even of the individual items. On the test form, the items in the test are arranged in order of difficulty, so that many of the informal numbering items appear at the beginning of the test while many of the calculation items are more toward the end of the test. The test is not timed.

4. Administration Procedures: The TEMA kit consists of an examiner's manual, a picture book, profile/examiner record booklets, and manipulatives (for a few of the items). As with EGRA, the assessor begins with a set of simple instructions. The assessor then fills in information about the child on the instrument and begins with the items. Each item has its own set of instructions. A unique feature of the TEMA is that there are different entry points depending on the age of the child. The youngest children (age 3) begin with item #1, but children of other ages begin at different points in the test. The assessor also establishes a basal (a lower bound) and a ceiling (an upper bound) for each child. Both the different starting points and the establishment of a basal and ceiling are designed to reduce the administration time while obtaining valid scores. On some of the items, there is a practice item so that the child has a better idea of what they are supposed to do.
5. Scoring: Each item is scored as either right or wrong. There are criteria for making scoring judgments. These criteria are explained in the examiner's manual, but are also noted on the score sheet. The assessor calculates a total score for each child.
6. Reporting: The TEMA allows for different types of score interpretations. In addition to a total score, there are tables for calculating grade level equivalents and a math "IQ" score, with a mean of 100 and a standard deviation of 15. These types of norm-referenced scores are perhaps more useful for children in high income countries, but they also provide a yardstick by which to judge the competencies of children in low income countries. In research conducted in Benin, for instance (Davis, 1992), the scales were useful in making cross-national comparisons.
7. Strengths: The main strength of the instrument is its basis in Piagetian and research-based developmental psychology of how young children learn mathematics. This includes testing both conceptual and procedural knowledge and skills. Another strength of the instrument is to provide information on both informal and formal mathematics of young children. These strengths are complemented by other features beyond those of other assessments, such as comprehensive information on validity and reliability, multiple score interpretations (e.g., scale scores, norm-referenced scores, grade level equivalents, math IQs), alternate equated forms, supplemental materials (e.g., a book on assessment probes and instructional activities), and DIF studies (differential item functioning, to check for item bias).
8. Weaknesses: Weaknesses of the instrument are related to its length. A typical administration to a child can take up to 30 or 40 minutes. To cut down on administration time, the TEMA developers created a way of setting a basal and a ceiling, if the child has correctly answered or incorrectly answered 5, then the basal or ceiling is set. This can be confusing to some test administrators, even after training. It is not curriculum based but does have strong ties to many curricula. It also loses many of its strengths when it is adapted, such as the scales.
9. Suggestions for Improvement: The TEMA has gone through substantial revisions and has answered many of the issues posed by critics. These include using manipulatives for various tasks, printing the picture book in an easel format, providing explicit descriptions of items, improving the recordkeeping systems, and the using linear test equating for parallel forms.
10. Countries: The TEMA instrument has been in print for the longest period of time and has been used in several countries outside of the U.S., including in Africa, Asia, the Caribbean, and Europe. It has been used extensively in research and program evaluation, though we did not find direct evidence of its contribution to policy dialogue in low income countries.

ASER/UWEZO

1. Needs Addressed: The ASER center in India, under Pratham, developed a household-based, locally administered national level assessment in basic literacy and numeracy. It is currently the only data source on elementary school children's learning levels in India. These assessments were developed out of a need to provide independent evidence of children's progress towards learning basic literacy and numeracy skills in India. UWEZO has adapted the ASER assessments for use in East Africa, and ASER tools are currently being used in Pakistan.
2. Test Development: ASER began to collect data in 2005 using internally developed tools. The tools are aligned to the Grade 1-4 curriculum, and are individually administered. The measures have been validated through several studies. The ASER study also includes information on infrastructure, school enrolment, and attendance.
3. Competencies and Items: The ASER assessment has used a core set of tasks from 2005 to 2011, which were given to all children between the ages of 5-16. There have been other tasks that were used infrequently. For example, in 2006, ASER used word problem items involving subtraction and division. In 2007 and 2008, there were word problems with currency and telling time. In 2010, there were calendar, area, and estimation tasks. The core tasks given every year are:
   - Number Recognition: 1-9
   - Number Recognition: 11-99
   - Subtraction: 2-digit numbers with borrowing
   - Division: 3-digit number by a 1-digit number with remainders
4. Administration Procedures: ASER recruits a large number of volunteers from NGOs, citizen groups, and educational and government institutions to assess children each year. ASER provides a short document explaining the assessment procedures. For example, the assessor is told to begin with the subtraction problem; if the child gets two incorrect responses in a row, the assessor moves to number recognition tasks 11-99.
5. Scoring: For each set of items (i.e., subtraction problems), the assessor determines if an answer is correct or incorrect, which then determines the subsequent problems given to the child. If a child solves a certain number of problems in a set correctly, he/she is scored as "a child that can do subtraction." If a child cannot solve a certain number of problems correctly, he/she is scored as "a child that cannot do subtraction."
6. Reporting: The assessor scores each assessment as it is being done. Then, community members gather and create a village report card, which has information on enrollment, attendance, and basic learning for all children in the village. Based on this information, next steps are planned by the community.
7. Strengths: The assessment tools are easy to use by volunteers with minimal training. The results are generated quickly and are easy to interpret, and can provide immediate feedback to teachers, parents, and other invested in education. Due to ease of use, manner of implementation, and low cost, ASER assessments are easily scalable (they currently assess about 700,000 children annually). Because ASER assessments are designed as a tool that can be used by anyone, it encourages parents/local community members to take responsibility for children's learning.
8. Weaknesses: The ASER tools measure largely procedural knowledge, without focusing on conceptual knowledge. The tools also focus largely on formal mathematics. There may also be problems with implementation due to scale and use of minimally trained volunteers to assess children.
9. Suggestions for Improvement: The ASER assessment model can be a powerful model that reaches a large number of children while at the same time encouraging community awareness to improve education. In terms of improvement, the tools may need to be modified to include conceptual understandings as well as informal mathematics.
10. Countries: The ASER/UWEZO instrument has been used in India, Pakistan, Kenya, Tanzania, and Uganda. In India, in particular, ASER reports are oriented toward not only improving classroom learning but also toward improving educational policy. For instance, in one of their reports, they present low results on literacy and numeracy assessments and then state that the policy makers should put more funding into better instruction and invest less in infrastructure.

B. Comparison of the assessments by content

This section addresses the content of the four assessments (ASER and UWEZO separately) by comparing them to each other according to the different competencies in the Common Core State Standards. These standards define the numeracy topics that should be taught in the U.S., and they are a reflection of the most up-to-date international research on early grade numeracy competencies.

For grades K-2, the Common Core State Standards are divided into three domains: 1) counting and cardinality; 2) operations and algebraic thinking; and, 3) number and operations in base-10. The domain of counting and cardinality includes using numbers and written numerals to represent quantity and solve problems. Operations and algebraic thinking includes using written symbols to represent problems, comparing numbers, and composing and decomposing numbers. Number and operations in base-10 is about place value, understanding the organization of the number system, and using this understanding to solve addition and subtraction problems. The domains are further divided into sub-domains that provide guidelines on what children should be able to do in grades K-2. The domains and sub-domains are listed in the table below.

Domains	Sub Domains	EGMA	TEMA	ASER	UWEZO
Counting and Cardinality	Count to tell number of objects		X		X
	Compare numbers	X	X		
Operations and Algebraic Thinking	Understand addition and subtraction	X	X		
	Solve problems with addition and subtraction	X	X	X	X
Number and Operations in Base-10	Read and write numerals	X	X	X	X
	Extend counting sequences	X	X		
	Understand place value		X		
	Use place value to add and subtract	X	X	X	X

Based on this list of competencies, we review which of these competencies current early grade numeracy assessments cover. It is important to keep in mind that, of course, not all of these competencies need to be measured by one assessment. It is also important to note that some competencies may be assessed in different ways (e.g., through classroom observations, through curriculum-based assessments) rather than through standardized diagnostic instruments.

The most comprehensive assessment is the TEMA, which is not surprising since it is twice as long as any of the others (i.e., about 30-40 minutes for the test administration) and is strictly based on current research in mathematics education. EGMA covers most of the Common Core subdomains, while ASER and UWEZO (the shortest of the assessments) each have content in fewer subdomains.

The EGMA test had counting in its first iteration, but it was eliminated in the second version due to children's high performance. Place value is not specifically addressed. The ASER and UWEZO tools only measure number and computation skills.

C. Need for further development of existing tools

While each of the tools described in this study has many advantages, there are needs for further development, which in fact is recognized by the authors of each of the tools. EGMA is now on its second version, with the elimination of oral counting and one-to-one correspondence (see above), and the addition of relational reasoning and spatial reasoning. The TEMA is currently in its third edition, with the addition of more informal mathematics items. ASER and UWEZO have different versions of their instrument. The EGMA, ASER, and USWEZO instruments are more flexible than TEMA, and the authors encourage adaptations to fit the needs of local contexts. The process of refinement for each of the core instruments is iterative, based on the current research in early numeracy.

One area that will likely appear more on future versions is higher-level cognitive skills, including more conceptual and problem solving skills. Another area of growth is informal mathematics, as in the TEMA. It is possible that future versions of EGRA, ASER, UWEZO might include more measures of informal learning; this is particularly the case if their developers would like for their instruments to be used with younger children, i.e., kindergarten or pre-school.

Another area of adaptation will be in the use of technology for administering the instruments. This will happen for several reasons. One, the technology (e.g., for tablets, netbooks, cell phones, and PDAs) is becoming less expensive each year and the devices have more capabilities. Two, the data entry becomes more straightforward and less burdensome if the children's responses can be entered electronically. Third, there are some items that can more easily be administered using technology. For instance, in the development of the EGMA, the expert panelists recommended including an item on number lines. However, the number line task recommended was generally administered using a computer. The panelists adapted the number line task for paper-and-pencil use, but it was eventually discarded due to difficulties in obtaining accurate scores using the paper-and-pencil version. These kinds of problems will likely have viable technology-based solutions, which can be applied in either high or low income countries, in the near future (see part E of this section).

A final area is expanding items to include situations such as those found in business, health, and other practical areas. The recent UNESCO/Brookings international standards had a focus on practical numeracy for primary school children. Items may need to be developed which assess children's understandings of the practical applications of numeracy, such as operations with currency.

D. Need for development of whole new tools

The existing tools cover the needs for diagnostic and summative mathematics assessments. While some of the tools need improvement, and others could benefit from more stable funding (e.g., the regional assessments), the basic instruments and procedures are in place. Most of the existing tools are somewhat flexible, so that the items can be changed while maintaining the basic structures of the instruments.

In contrast, the other main category of assessment – formative (classroom) assessment – is the least discussed and has the strongest need for more development. The remainder of this subsection is devoted to providing more information on formative assessment, including evidence, implementation, and policy implications.

Formative assessment

Formative assessment (also called school-based, classroom, and continuous assessment) has high potential for raising achievement levels in early mathematics. Black and Wiliam (1998) concluded that there is "no other way of raising standards (achievement) for which such a strong prima facie case can be made." Kellaghan and Greaney (2003) wrote that while formative assessment has "attracted the least attention in proposals to use assessment to improve the quality of education, it is likely to have a greater impact on student learning than any other form of assessment."

The reason for formative assessment's potential is that it is at the heart of effective teaching and improved student learning. Achievement is driven primarily by what teachers and students do in the classroom. The tasks of the teacher are complicated and demanding – they involve managing 30 or more children, delivering instruction, and addressing individual needs. Learning outcomes in numeracy can be improved only if the teacher is able to gather information on student behavior and use that information to organize daily tasks and deliver instruction more effectively.

In addition to using formative assessment information to improve learning, there is a concern about the fairness of evaluating students by only using summative assessments. In extreme situations, particularly in low income countries, students are only evaluated once per year in mathematics through a summative examination. Even if students are also assessed at the end of each term, or perhaps three to four times per year, it seems fairer that students should be evaluated at multiple points per term, even if the assessments are short or informal, so that they can demonstrate their learning more frequently and in multiple ways (Nitko, 1994).

There are three main techniques used in formative assessment of early numeracy: observation, questioning, and tasks. All techniques can be used with individuals, groups, or entire classes. First, observation should be based on prior knowledge about the mathematical activities that students are supposed to perform. The teacher can make notes about the performance of children. Next, the teacher should pose questions to the student based on observations. As Piaget (1976) said, questioning keeps us from "restricting ourselves to observing the child without talking to him." Finally, the teacher may use tasks in conjunction with observation or questioning, or at the end of the lesson as a part of classwork or homework. The same task given to students can tell the teacher about differences in student performance. Teachers can also follow up by asking targeted questions of specific students. For example, with an addition task, the teacher can ask questions such as "How did you know that was the answer?" or "Can you show me how you solved the problem?"

Most formative assessments are informal, consisting of observing and questioning students as they attempt to perform mathematics. Informal formative assessments are complimented by formal formative assessments, which are usually in paper-and-pencil format, e.g., at the end of a curriculum unit. Formal formative assessment an also include reviewing classwork and homework, as well as administering quizzes.

In all of these methods – observing, questioning, and tasks – the basic premise behind formative assessment is that teaching and learning must be interactive. In early numeracy instruction, Ginsburg (2009) stated that the "bottom line attribute of formative assessment is its actionable character: it is assessment that informs instruction in an effective manner." It cannot be overly academic and finely detailed, but at the same time cannot be too broad and general either. As Ginsburg noted, it must be mid-level, i.e., based on some knowledge by the teacher on how children learn mathematics and oriented towards specific pedagogical interventions to help children learn better.

Finally, formative assessment must be a part of a systematic approach. McKinsey (2010) researchers conducted a study of early literacy and numeracy achievement in five countries: Brazil, Ghana, India, Jordan, and South Africa. They concluded that improvements in learning outcomes require a formative assessment system that 1) sets minimum proficiency targets for schools and students, 2) conducts frequent student learning assessments (every 3-4 weeks), and 3) processes and uses data to monitor school and student progress.

Evidence

How do we know that formative assessment can improve student performance? Black and Wiliam (2001) conducted the most extensive study on formative assessment to date. They reviewed 580 articles and chapters on various programs and interventions involving formative assessment. In their research, they asked three questions:

1. Is there evidence that formative assessment raises learning?
2. What are some of the current problems with formative assessment?
3. How can we improve formative assessment?

On the first question, the authors examined the effect sizes – a statistical indicator of the size of improvements from an intervention – gathered from evaluations of formative assessment programs. The effect sizes typically ranged from 0.4 to 0.7 (in standard deviation units). Such effects translate into improvements of the average performance of students by between one and two grade levels. In addition, formative assessment was found to generally benefit low achieving students more than high achieving students, thus helping to reduce the gap in achievement levels.

On the second question, problems found in the literature with current formative assessment practices can be summarized as follows:

- Many teachers simply mark papers and provide scores to students, without analyzing results and adjusting their instruction.
- Teachers often use marks and scores to put students in competition with each other rather than focusing on improvements for all students.
- Teachers tend to encourage rote and superficial learning, which is easier to assess, even when they say that they want to develop understanding of concepts.
- Teachers often emphasize quantity of learning of a wide range of competencies at the expense of quality of learning of the most important competencies.
- Teachers usually do not share and discuss their assessment methods and results with other teachers.

On the third question, the following main points are backed up with evidence on how to improve formative assessment:

- Teachers must change their belief structures about students in two fundamental ways. First, many teachers believe that knowledge is simply to be transmitted and it is up to the students to learn. However, more and more teachers, including those in low income countries, are realizing that the transmission model does not work, but interactive teaching does. In early numeracy, what is needed is a culture of observation and questioning about concepts and skills (Ginsburg, 2009). Second, there is the belief by many teachers that only certain students can learn. Again, more and more teachers are realizing that all students can become proficient in literacy and numeracy, if we can overcome the obstacles to learning (McKinsey, 2007).
- Teachers must provide relevant feedback and corrective strategies to students on a frequent and regular basis. This needs to happen within a culture of success, i.e., in an environment where students are encouraged to improve and succeed. Students can work with feedback pointing out that they are not grasping the material if they are not put in a situation where the communications are negative, punitive, or overly oriented towards competition with other students. Studies have shown that students only benefit marginally from feedback in the form of marks or grades; they also need explanations and strategies to help them learn better (Black & Wiliam, 1998). As Ginsburg (2009) stated, teachers of early numeracy need to give students information about what is wrong and make sure that they have approaches on how to put it right. He focuses extensively on "debugging" procedures in solving problems.
- Teachers need help to implement formative assessment. In order to make formative assessment work, teachers must have support in the form of training, supervision, and on-going mentoring. Many education reforms concentrate on inputs – new curricula and materials, management rules and requirements, testing programs – with the expectation that outputs will naturally follow – higher student scores, more students progressing through the system. However, this seldom works. It is unfair to leave the most difficult piece of the learning improvement program – the teaching and learning in the classroom – entirely to the devices of the teachers, particularly those in low income countries who are often situated in isolated rural schools.
- Teachers can succeed if they receive help and support. The evidence shows that teachers can successfully implement formative assessment if they receive proper training and support. As stated by Ginsburg (2009), "a key challenge for mathematics education is to provide professional development designed to help teachers to understand and assess children's mathematical minds." There is no question that it is hard work, but, if it is done effectively, positive results are almost sure to follow.

Implementation

The process of designing and implementing a formative assessment program, and subsequent changes in instruction, is slow and takes place through sustained, practical professional development and support. This is the only way for fundamental improvements in teaching and learning to occur.

There are five steps in the process. These are general steps since the actual program depends on previous formative assessment programs and the context of schooling within a country or region.

1. Develop examples of formative assessment in the classroom that teachers can learn from and use in their practice. Teachers will not change as a result of simple "chalk-and-talk" training or by receiving a few materials. Formative assessment needs to be tried out in a small number of schools and practical examples – with videos and/or personal testimonies – need to be drawn from those schools and teachers.
2. Use the examples in a hands-on training program led by master trainers. A group of trainers needs to actively share the work from the pilot schools. The dissemination will be made more effective by using videos and actual classrooms of children. Teachers need living examples of implementation, as practiced by teachers with whom they can identify and from whom they can derive the confidence that the program will work.
3. Provide print materials to teachers. These should be in the form of an easy-to-use tool kit that is linked to the curriculum and lesson plans. The kit should provide questioning and observation techniques for assessing children during lessons and tasks to assign for classwork, homework, and quizzes. Teachers should have guidance from outside sources, including print materials on implementing formative assessment in their classrooms.
4. Reduce obstacles to implementation. Teachers need to combine the guidance from outside sources with the space to figure out how to make the program work in their own classrooms. They must also have the opportunity to collaborate with other teachers, both in their own schools and in neighboring schools, so that they can share ideas and methods. This can be accomplished through "lesson study" (Fernandez, 2002) or "communities of learners" (Bransford, Brown & Cocking, 1998). In addition, specialists should visit the classrooms to provide supervision and support. Teachers also need to know that formative assessment information is not a reflection of how well they are doing their job and will not be used to evaluate them.
5. Use formative assessment results to improve instruction. Teachers can develop learner profiles – or performance descriptors – to provide specific support for groups of students and individual learners. They can use this information to organize their instruction, using ideas in teachers guides and other materials, and also from collaborating with other teachers, in an extension of the "lesson study" and "communities of learners" structures.

The role of the teacher is to facilitate improvement for all students in early numeracy. An evidence-based way to accomplish this is by using assessment information to develop specific instructional activities for the class as a whole and for individual or groups of students. For instance, a child may understand addition but needs practice to memorize number facts so that the addition process is automatized. Another child may need help in developing concepts for explaining their procedural knowledge in subtraction. Formative assessment can help the teacher to identify thinking processes and obstacles, which can then be used in developing pedagogical approaches.

Policy

A final point is in the area of formative assessment policy. In the past, most assessment policies have concentrated on testing the students and establishing a competition for promotion, selection, and certification. This has resulted in norm-referenced assessments in which students either pass or fail their tests. Current policies are shifting towards setting learning targets for all students, with assessment used to provide a check on student attainment of the targets. But this needs to go further. Feedback from assessments should be communicated directly to students, along with strategies to help them master numeracy content. Most of this feedback should take place during formative assessment in the classroom.

The overarching policy priority in assessing learning outcomes in mathematics, and in other subject areas, should be the promotion and support of more effective learning in the classroom. Policies should focus on proven strategies for improving teaching and learning: 1) interactive teacher-student communications, 2) targeted instruction and support for students based on objectives, and 3) particular help for lower achieving students.

Black and Wiliam (2001) state that governments, school authorities, donor agencies, and the teaching profession need to think very carefully about whether they are serious in raising learning outcomes. This is because change occurs slowly and there is no magic bullet, particularly when it comes to teacher behavior in the classroom. However, if there is a strong level of commitment on the part of all stakeholders, strategies can be developed so that the evidence-based formative assessment approaches can be acted upon, and learning can improve, particularly in the core areas of literacy and numeracy.

E. Opportunities and challenges of using mobile technologies for assessment

There is tremendous potential in using mobile technologies for assessment. Until recently, many of the assessment applications in low income countries focused on using laptop computers. However, the popularity of tablet computers could have a revolutionizing effect on assessment, particularly if the assessment results can be uploaded via Wi-Fi or cell phone networks. Smart phones and personal data assistants (PDAs) have equal potential for use as mobile technologies for assessment.

Mobile devices have two main advantages: access to information and learner engagement. Smart phones are perhaps the wave of the future in terms of access and engagement. There is already 20 percent penetration by smart phones in low income countries (as opposed to 50 percent in high income countries).

While penetration is relatively easy to measure, little evidence is available on engagement, particularly in regards to mobile learning. Few examples exist of using mobile devices to improve school learning. However, new applications are being developed on a constant basis, so we can expect that mobile devices will become a part of classroom pedagogy. In addition, children learn to navigate mobile devices very quickly and become engaged almost immediately in many cases.

There is some evidence of game-based learning using mobile devices in the U.S., India, and Peru. Some of the lessons learned include the importance of adapting the content to specific environments, moderate effects on achievement, and little change in student motivation. However, these results will likely improve as applications become more aligned to the needs of the classroom and the psychology of the children.

In the area of early numeracy, applications have been developed for young children in writing and tracing numerals, counting, comparing, matching, and identifying shapes. There are also ways of transferring information from the teacher to the student and vice versa, or to and from the parent. Examples are the following:

- Tablet Math (U.S.)
- Text2Teach (Global)
- Math4Mobiles (Israel)
- Dynamic Protractor (U.S.)
- Dr. Math (South Africa)
- MoMaths (South Africa)
- Splash Math (U.S.)

For more information, the topic of using mobile technologies for learning and assessment is addressed in the companion paper to this study on learning outcomes (Stringel & Pouezevara, 2012).

Summary

The three widely known diagnostic assessments – EGMA, TEMA, and ASER/UWEZO – have been developed to meet specific needs in numeracy outcomes. Information has been presented in this section so that the user can make decisions on which assessment fits a particular context. The assessments can be adapted to varying degrees for further customization.

A particular focus was placed on formative assessment as perhaps the best means for increasing student learning through the use of assessment. A quick guide was presented to developing and implementing a formative assessment program.

This section concluded with some brief comments on mobile learning and numeracy, including a reference to a companion paper that explores this vital topic in further depth. Mobile learning has potential not only in the area of assessment but also in improving instruction and learning in early numeracy.

Question 4: What can be learned from the assessment of early grade literacy with reference to numeracy?

As mentioned in the terms of reference for this desk study, the area of literacy is ahead of numeracy in assessing learning outcomes in low income countries. While the traditional assessments – public examinations, national assessments, international assessments, and classroom-based (formative) assessments – have maintained a balance between literacy (language) and numeracy (mathematics), the emphasis on diagnostic assessment is mostly in the area of early literacy.

While it is understandable that literacy would be in the forefront of efforts both in instruction and assessment given the strikingly low levels of literacy in low income countries, numeracy is an equally critical component of schooling. A better approach would be to ensure that both literacy and numeracy receive the attention they deserve. Each is necessary in young children's schooling but neither is sufficient. In addition, studies have shown that they reinforce each other. This section provides background on the complementarity of literacy and numeracy, the differences between literacy and numeracy, lessons learned from the promotion of literacy that can help in numeracy, and the importance of numeracy in predicting later achievement in literacy.

A. Numeracy and literacy as complementary

A review of numeracy and literacy leads to several basic conclusions. First, children must become competent in both domains. Having strong knowledge and skills in either numeracy or literacy is not enough for later educational and occupational success. As stated by the Mathematics Learning Study Committee (2001),

"To be employable in the modern economy, high school graduates need to be more than merely literate. They must be able to read challenging material, to perform sophisticated computations, and to solve problems independently."

Second, the foundation for both numeracy and literacy is developed early. Learning to read and developing mathematical proficiency both rest on a foundation of concepts and skills that are acquired by many children by grade one. In the case of reading, children are expected to enter school with a basic understanding of sound structure, an awareness of the alphabetic writing system, and skill in handling basic language concepts. Likewise in mathematics, students should possess a toolkit of basic mathematical concepts and skills when they enter first grade. This includes comparing quantities, counting, knowing the concepts of addition and subtraction, and recognizing simple patterns and shapes.

Third, in both domains, there is evidence that early intervention can prevent full-blown problems in school. If children have not mastered certain basic skills, they can expect problems throughout their schooling and later. Research on reading indicates that all but a very small number of children can learn to read proficiently, though they may learn at different rates and may require different amounts and types of instructional support. Similar observations can be made for numeracy.

Fourth, literacy and numeracy merge in certain early grades activities. For example, nearly all grade two children can be expected to make a useful drawing of the situation portrayed in an arithmetic word problem as a step toward solving it. Representing numbers by means of a drawing is a task that children should be able to do in early primary school. In addition, children at grade two should be able to learn the basic symbolic or abstract elements of literacy. The fact that letters represent sounds and that combinations of letters make other sounds is an early skill. The Hindu-Arabic numerals used in everyday life are also symbolic in that a one-digit number represents a certain quantity. Of course, this is where letters and numbers diverge (see the next subsection below).

Fifth, reading specialists have developed a variety of programs for students having reading difficulties so that prompt and effective assistance is available. Similarly, mathematics programs can help children who need help with basic concepts. These programs focus on foundational skills that children bring from outside of the school environment and build on the formal knowledge and skills that are necessary for school success.

B. Ways that literacy and numeracy are different

There are distinctive differences in literacy and numeracy for young children, particularly those who are learning to read and understand math. One of the main differences is in the symbols. As mentioned in the previous subsection, letters and simple numbers have some commonalities in that they are representations of sounds and quantities, respectively. In English, both letters and numbers are members of a finite set (26 letters and 10 digits). Both are used in combination with other symbols, such as punctuation in reading and operational and relational symbols in math. Both have names that have nothing to do with the symbol's actual meaning or value.

However, the symbols have strong distinctions as the numbers become more complicated. In most languages, the alphabetic writing system, once learned, enables the student to read and decode any word or sentence, although of course not necessarily to understand its meaning. Mathematics, in contrast, has many types and levels of representation. In fact, mathematics can be said to be about levels of representation, which build on one another. One of the most important ideas with numbers is that they can be composed and decomposed. Letters cannot be broken apart. We can break down larger numbers into smaller numbers, for instance the idea that five is the same as two and three together. Standalone letters do not have such a purpose. It is only when letters are put together that they have a purpose. From letters, we can build sounds and words, and then sentences, paragraphs, and so on. On the other hand, when we put together two numbers (or three or more numbers) we are only left with another number. Importantly, numbers can take on different meaning in different contexts. With place value, for instance the 7 in 17 means seven ones, while the 7 in 71 means seven tens. Letters have very limited variation (long a, short a, etc.) but for numbers the transitions are limitless.

It is important for children to recognize and understand these differences between letters and numbers, and between literacy and numeracy. Understanding the distinct qualities of numbers will allow children to make more sense of numeracy.

C. Building on knowledge/ experiences from literacy to promote numeracy

The main literacy proponents in the donor world – GPE, USAID, World Bank, EU, DfID, AusAID, and others – have been successful in publicizing the low levels of literacy in low income countries. They are also currently experiencing increased success in orienting development projects towards literacy interventions. This should lead to better rates of literacy among young children over the next several years.

Similar work could be (and should be) accomplished in numeracy. The GPE in particular has drawn some lessons learned from the literacy movement that can be applied to numeracy, and perhaps will be able to shortcut some of the problems that have hindered more rapid progress in literacy (Crouch, 2012).

One of the early realizations in the promotion of literacy several years ago was that many people were aware of the problem of low levels of literacy but there were no accepted ways of measuring it. In addition, development experts were not exchanging information in an organized manner and there were few successful experiences of successful change. Donor agencies did not have goals that centered on improved children's reading. Currently, however, much of this has been reversed. As described in the earlier sections of this study, there are several measures of early literacy that are being implemented in low income countries. Donors are concentrating funding in literacy and they are coordinating their efforts in some countries. The area that is most in need of greater progress at this time is that there are not enough examples of large-scale improvements in children's reading levels.

The main lesson drawn from the GPE in making progress in literacy is partnerships. There are several elements in literacy partnerships that seem to be working. One element is the need for a community of practice (CoP) that can develop and promote standards for literacy programs. The CoP is currently promoting and funding the measurement, assessment, and goals for children's early literacy. It is gathering and managing experiences that show impact. It is also promoting literacy through the social marketing of key literacy statistics and indicators.

In the area of social marketing, the literacy CoP is using oral fluency (i.e., most grade two children cannot read a single word in many countries) as an indicator that can galvanize stakeholders. However, the numeracy CoP has not yet identified a similar indicator. In addition, the literacy CoP has been instrumental in pushing for

measureable goals, such as cutting in half the number of non-readers in Grade 2 in 5 years, or helping 100 million children improve their reading by 2015.

To develop more powerful indicators, there is a need to link oral assessments of early reading and math with written assessments, both within countries and internationally. This will require good practice standards so that the linking is done in a professional way. Issues to overcome are the periodicity of the regional and international assessments, i.e., every three or four years, and the need to support the regional assessments, which need both financial and managerial assistance.

All of this requires action, such as conferences involving governments, non-governmental organizations, donor agencies, and other champions. Much of it involves measuring children's numeracy outcomes, and using the information as a critical ingredient in formulating goals and indicators. It also requires the formal knowledge management of the assessment results and their linkages to the goals and indicators, as well as to children's later achievement in school.

D. Using early numeracy to predict later achievement

It is often thought that early reading skills are the best predictor of later reading and math achievement. However, recent research shows that this is not the case. In a widely cited study, Duncan et al. (2007) used large longitudinal databases from the Canada, the U.K., and the U.S. to estimate the effects of school readiness – in reading, math, attention skills, and social skills – on later school reading and math achievement. Across all databases, the strongest predictor of later achievement was school-entry math, followed by reading and attention skills. The measures of social skills had no significant effects on later achievement. Patterns of these associations were similar for boys and girls and for children from high and low socio-economic backgrounds.

The following table shows the standardized regression coefficients for the predictors.

Meta-Analytic Regression Results (Standardized Coefficients)

Independent Variables (School-Entry Meassures)	Outcome Variables (Later Achievement)		
	Reading	Math	Reading & Math
Reading	0.24*	0.10*	0.17*
Math	0.26*	0.42	0.34*
Attention skills	0.08*	0.11*	0.10*
Social skills	-0.00	-0.01	-0.01

* Statistically significant at $p < 0.001$

The coefficients show the strength of the independent variables (school-entry reading, math, attention skills, and social skills) on the outcome variables (later achievement in reading and math, along with a combined measure of reading and math). As hypothesized, school-entry reading is a strong predictor of later reading achievement (0.24) and school-entry math is a strong predictor of later math achievement (0.42). However, contrary to most of the literature, school-entry reading was not a strong predictor of later achievement in math. Most surprisingly, school-entry math was a strong predictor of later achievement in reading, and in fact a stronger predictor than school-entry reading. On the combined indicator of later achievement in reading and math, school-entry math was twice as strong of a predictor as school-entry reading (0.34 vs. 0.17).

There were two other findings of note from the Duncan study. One was the importance of measuring children's early literacy and numeracy skills. As stated by the authors:

> *"All of our data sets suggest that reading and math tests that were individually administered to children by trained personnel around the point of school entry can be a highly reliable way of assessing early skills."*

In addition, the authors stated that:

> *"It was also the case that we could not attribute most of the variation in later school achievement to our collection of school-entry factors, so the potential for productive interventions during the early school grades remains."*

In particular, they mentioned the potential of curricula designed with the developmental needs of children in mind, especially those that are engaging and fun. An example cited was the Big Math for Little Kids program (Greenes, Ginsburg, & Balfanz, 2004), which capitalizes on children's interest in exploring and manipulating numbers.

Several researchers have replicated and expanded upon the Duncan study (Foster, 2010; Grissmer et al., 2010; Pagani et al., 2010; Romano et al., 2010).

Summary

Linkages between literacy and numeracy were examined in this section of the study. Some of the similarities and differences between literacy and numeracy were addressed. The main point was to recognize these variations so that assessments (and interventions) would take them into consideration.

Recent lessons learned from international experiences in promoting young children's literacy were also discussed. The hope is that efforts to improve children's numeracy will be able to draw on the experiences from literacy.

Finally, a statistical study using early literacy and numeracy to predict later school achievement was described. The conclusion from the study was that early numeracy was a stronger predictor of later reading and math than early literacy. The researchers also noted that measuring young children's literacy and numeracy was a highly reliable way of gathering information on early knowledge and skills.

Final Comments

The desk study has provided an important opportunity to address the four questions posed by GIZ:

1. What kinds of numeracy learning outcomes need to be assessed?
2. What are the current practices in assessment of early grade numeracy learning outcomes?
3. What (low-cost, effective, and easily applicable) early grade numeracy assessment tools are available?
4. What can be learned from the assessment of early grade literacy with reference to numeracy?

While many assessments exist to measure early numeracy, the most important part of assessment is to define the purpose. We need to determine whether information is needed for program evaluation, policy dialogue, or classroom instruction. The other critical part is that the assessment needs to be done well, so that standards for reliability and validity are respected and that the information is used for the intended purpose.

Finally, in the TIMSS 2011 report (Mullis et al., 2012), the authors highlighted the importance of an early start in shaping children's numeracy skills. The TIMSS assessment results at grades 4 and 8 indicated that students had higher performance in mathematics if their parents reported that:

- They often engaged in early numeracy activities with their children;
- Their children had attended pre-primary education; and
- Their children started school able to do early numeracy tasks (e.g., simple addition and subtraction).

The authors also emphasized the importance of combining early literacy and numeracy during the preschool and early school years. They also mentioned early numeracy activities as significant predictors of later mathematics scores. Parents who played games and engaged in other numeracy-related activities with their children saw the rewards from their efforts; scores for children benefiting from early numeracy scored an average of 50 scale score points higher than children who were almost never involved in early activities. Similarly, children who could perform early numeracy tasks at school entry scored an average of 73 scale score points higher than children who did not possess such abilities.

These kinds of assessment results can be instrumental in providing background information that can lead to successful interventions and improved numeracy learning outcomes. The keys are to collect useful assessment information and then ensure that it is applied towards better learning for children.

References

Anderson, K., Case, A., & Lam, D. (2001). *Causes and consequences of schooling outcomes in South Africa: Evidence from survey data.* Ann Arbor, MI: Population Studies Center.

Baranes, R., Perry, M., & Stigler, J.W. (1989). *Activation of real-world knowledge in the solution of word problems.* Cognition and Instruction, 6(4). 287-318.

Baroody, A.J. (1987). *Children's mathematical thinking.* New York: Teachers College Press.

Black, P., & Wiliam, D. (1998). *Assessment and classroom learning.* Assessment in Education, 5(1), 7-71.

Black, P., & Wiliam, D. (2001). *Inside the black box: Raising standards through classroom assessment.* London: Assessment Group of the British Educational Research Association.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.) (1998). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Carpenter, T., Fennema, E., & Franke, M. (1996). *Cognitively guided instruction: A knowledge base for reform in primary mathematics.* The Elementary School Journal, 97(1), 3-20.

Carraher, T.N. Carraher, D.W. & Schliemann, A.D. (1985). *Mathematics in the streets and in schools.* British Journal of Developmental Psychology, 3, 21-29.

Clements, D.H. (1999). *Subitizing: What is it? Why teach it?* Teaching Children Mathematics, 5(7), 400-405.

Crouch, L. (2012). *Literacy to numeracy: Lessons in creating a movement?* Berlin: BMZ GIZ GPE Numeracy Conference.

Davis, J.C. (1992). *Young children's mathematical understanding in Benin and the United States.* Unpublished doctoral dissertation. New York: Teachers College, Columbia University.

Davis, J.C. & Ginsburg, H.P. (1993). *Similarities and differences in the formal and informal mathematical cognition of African, American, and Asian children: The roles of schooling and social class. In Cognition and culture: A cross-cultural approach to cognitive psychology.* Amsterdam, Netherlands: Elsevier Science Press.

Duncan, G. et al. (2007). *School readiness and later achievement.* Developmental Psychology, 43(6), 1428-1446.

Dweck, C.S. (2006). *Mindset: The new psychology of success.* New York: Random House.

Fernandez, C. (2002). *Learning from Japanese approaches to professional development: The case of lesson study.* Journal of Teacher Education, 53, 393-405.

Foster, E.M. (2010). *The value of reanalysis and replication: Introduction to special section.* Developmental Psychology, 46(5), 973–975.

Gelman, R. & Gallistel, C. R. (1978). *What develops and how. In The child's understanding of number (pp. 202-242).* Cambridge, MA: Harvard, University Press.

Ginsburg, H.P. (2009). *The challenge of formative assessment: Children's minds, teachers' minds.* Human Development, 52, 109-128.

Ginsburg, H. & Opper, S. (1988). *Piaget's theory of intellectual development.* New Jersey: Prentice Hall.

Ginsburg, H. P., Klein, A., & Starkey, P. (1998). *The development of children's mathematical thinking: Connecting research with practice.* In W.V.B. Damon (Ed.), Handbook of Child Psychology, 4, 404-427.

Ginsburg, H. P., & Baroody, A. J. (1983). *The Test of Early Mathematics Ability.* Austin, TX: Pro-Ed.

Ginsburg, H. P., & Baroody, A. J. (2003). *The Test of Early Mathematics Ability (3rd ed.).* Austin, TX: Pro-Ed.

Gordon, P. (2004). *Numerical cognition without words: Evidence from Amazonia.* Science, 206, 496-499.

Grissmer, D. et al. (2010). *Fine motor skills and early comprehension of the world: Two new school readiness indicators.* Developmental Psychology, 46(5), 1008–1017.

Grogan, L. (2008). *Universal primary education and school entry in Uganda.* Journal of African Economics, 18(2), 183-211.

Guberman, S. (1996). *The development of everyday mathematics in Brazilian children with limited formal education.* Child Development, 67(4), 1609-1623.

Hatano, G. (1988). *Social and motivational bases for mathematical understanding.* In G.B. Saxe & M. Gearhart (Eds.), Children's mathematics. San Francisco: Jossey-Bass.

Kellaghan, T., & Greaney, V. (2003). *Monitoring performance: Assessment and examinations in Africa.* Paris: Association for the Development of Education in Africa.

Kellaghan, T., & Greaney, V. (2004). *Assessing student learning in Africa.* Washington, DC: World Bank.

Mathematics Learning Study Committee (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: The National Academies Press.

McKinsey (2007). *How the world's best-performing school systems come out on top.* London: McKinsey & Company.

McKinsey (2010). *How the world's most improved school systems keep getting better.* London: McKinsey & Company.

Miura, I., Okamoto, Y., Chungsoon, K., Steere, M., and Fayol, M. (1993). *First graders' cognitive representation of number and understanding of place value: Cross-national comparisons - France, Japan, Korea, Sweden, and the United States.* Journal of Educational Psychology 85(1), 24-30.

Mullis, I.V.S., Martin, M.O., Foy, P., and Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS Study Center, Boston College.

Nasir, N. (2000). *"Points ain't everything": Emergent goals and average and percent understandings in the play of basketball of African American students.* Anthropology and Education Quarterly 31(3), 238-305.

National Governors Association Center for Best Practices (2010). *Common Core State Standards for Mathematics.* Washington, DC: NGACBP.

Nitko, A.J. (1994). *Curriculum-based criterion-referenced continuous assessment: A framework for the concepts and procedures of using continuous assessment for formative and summative assessment of student learning.* Pretoria, South Africa: International Meeting of the Association for the Study of Educational Evaluation.

Nunes, T., Schliemann, A.D., & Carraher, D. (1993). *Street mathematics and school mathematics.* London: Cambridge University Press.

Pagani, L.S. et al. (2010). *School readiness and later achievement: A French Canadian replication and extension.* Developmental Psychology, 46(5), 984–994.

Piaget, J. (1964). *Development and learning.* In R.E. Ripple & V.N. Rockcastle (Eds.), Piaget rediscovered (pp. 7-20). Ithaca, NY: Cornell University.

Piaget, J. (1965). *Conservation of discontinuous quantities and its relation to one-to-one correspondence.* In The child's conception of number (pp. 25-38). New York: W.W. Norton & Company.

Piaget, J. (1976). *The child's conception of the world.* Totowa: Littlefield, Adams, & Co.

Romano, E. et al. (2010). *School readiness and later achievement: Replication and extension using a nationwide Canadian survey.* Developmental Psychology, 46(5), 995–1007.

Saxe, G. (1991). *Culture and cognitive development: Studies in mathematical understanding.* NJ: Lawrence Erlbaum Associates.

Saxe, G.B., & Posner, J.K. (1983). *The development of numerical cognition: Cross-cultural perspectives.* In H. P. Ginsburg (Ed.), The development of mathematical thinking. Rochester, NY: Academic Press, 291-317.

Sitabkhan, Y. (2009). *The use of convenient values among young train vendors in Mumbai, India. Proceedings of the Episteme 3 Conference, (pp. 114-118).* Mumbai, India: Homi Bhabha Center for Science Education, TIFR.

Snow, C.E., Burns, M.S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Stevenson, H.W., Lee, S.L., & Stigler, J.W. (1986). *Mathematics achievement of Chinese, Japanese, and American children.* Science, 231, 693-699.

Stigler, J. W., & Fernandez, C. (1995). *Learning mathematics from classroom instruction: Cross-cultural and experimental perspectives.* In C. Nelson (Ed.), Contemporary perspectives on learning and development: Twenty-Seventh Minnesota Symposium on Child Psychology (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates.

Strigel, C. & Pouezevara, S. (2012). *Mobile learning and numeracy: Filling gaps and expanding opportunities for early grade learning in low income countries.* Berlin: BMZ GIZ GPE Numeracy Conference.

Taylor, E. (2009). *The purchasing practice of low-income students: The relationship to mathematical development.* The Journal of the Learning Sciences,18(3), 370-415.

Wu, H. (1999). *Basic skills versus conceptual understanding: A bogus dichotomy in mathematics education.* American Educator/American Federation of Teachers, 1-7.

Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn,
Germany

Sector Programme Numeracy
Education Section
Division of Education, Health and Social Protection

Godesberger Allee 119
53175 Bonn
Germany
Tel. +49 (0) 228 44 60 - 0
Fax +49 (0) 228 44 60 - 1766

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn
Germany
Tel. +49 (0) 6196 79 - 0
Fax +49 (0) 6196 79 - 1115

numeracy@giz.de
http://www.giz.de/numeracy

In cooperation with
Jeff Davis
Yasmin Sitabkhan
School-to-School International

Design and Layout
Diamond media GmbH, Neunkirchen-Seelscheid
Cheryl Juhasz

Printed by
Top Kopie GmbH, Frankfurt/Main, Germany
Printed on FSC-certified paper

As at
December 2012

GIZ is responsible for the content of this publication.

On behalf of
German Federal Ministry for Economic Cooperation and Development (BMZ)
Education Division

Adresses of the BMZ offices

BMZ Bonn
Dahlmannstraße 4
53113 Bonn
Germany
Tel. +49 (0) 228 99 535 - 0
Fax +49 (0) 228 99 535 - 3500

BMZ Berlin
Stresemannstraße 94
10963 Berlin
Germany
Tel. +49 (0) 30 18 535 - 0
Fax +49 (0) 30 18 535 - 2501

poststelle@bmz.bund.de
www.bmz.de