Needs Assessment Analysis of Zambia's Basic Competence Testing Program: Observations and Recommendations

Diane R. Garavaglia, Ph.D.

May 2002

Contract GS-10F-0112J
Task Order No. 690-M-00-02-00001-00
EMIS and Related Activities
American Institutes for Research
USAID/Zambia
SO2: Improved quality of basic education for more school-aged children
IR 2.3: Improved information for education decision making processes

This document was created using Adobe Acrobat 6.0 Professional

Submitted by:
American Institutes for Research
1000 Thomas Jefferson St. NW
Washington, DC 20007
Phone: 202.403.5000/Fax: 202.403.5979
www.air.org

AMERICAN INSTITUTES FOR RESEARCH

Education Management Information System Project (EMIS)

Needs Assessment Analysis of Zambia's Basic Competence Testing Program: Observations and Recommendations

May 2002

Submitted by:
American Institutes for Research

Prepared by:
Diane R. Garavaglia, Ph.D.

Funded by:
United States Agency for International Development (USAID)

Presented to:
Examinations Council of Zambia (ECZ)

Introduction

The American Institutes for Research (AIR) is providing technical support to the Examinations Council of Zambia (ECZ) regarding its Basic Competence Testing (BCT) program at grades 4, 7, 9, and 12. AIR met with ECZ in Lusaka, Zambia (April 16 – 26, 2002) to gain an understanding of the current testing program; evaluate the current status of the testing program; and provide ECZ, the Ministry of Education, and USAID with observations and recommendations.

The visit in Zambia allowed AIR to work collaboratively with ECZ to support, enhance, and advise about ways to change its current BCT program, specifically at grade 4. (The BCTs also are administered in grades 7, 9, and 12 however they have not been fully implemented to adequately evaluate them.) The visit also allowed for the review of the Interactive Radio Instruction (IRI) program, which is being implemented in grades 2 and 3.

IRI is a method of teaching literacy and numeracy to students through radio broadcasts. The radio teacher and the classroom mentor implement the program. The daily lessons that the radio teacher and the mentors use are in the *Learning at Taonga Market Mentor's Guide.* The *Mentor's Guide* is published by the Educational Broadcasting Services and was developed using the same content objectives on which the BCT in literacy and numeracy are based. (See the *Learning Achievement at the Grade Two (2) Level in the Interactive Radio Instruction (IRI) Centres* for a more detailed description of the IRI program.) The students participating in the IRI program are tested to evaluate the degree to which they are learning the literacy and numeracy content objectives taught via the radio programs. The IRI uses examinations that are designed, developed, and administered by ECZ. Therefore, the visit to Lusaka afforded the opportunity to

assess whether the current testing system is a valid mechanism for assessing the impact of the IRI program on students' learning of basic competencies.

The intended outcomes of the visit included the Ministry of Education and ECZ to benefit directly from Dr. Garavaglia's local involvement with the project under Task 3.

The specific objectives for the first visit were to:

- Establish "teacher friendly" performance levels that reflect the curriculum and that enable reporting of results to pupils and parents.
- Assess the current tests in grades 2, 3, and 4 for their reliability and validity based on the curriculum.
- Evaluate the test development process.
- Provide best practices related to test administration in Interactive Radio Instruction cases for mentors and identify limitations and implications of using the radio in teaching and testing.
- Develop a monitoring mechanism to evaluate the use of tests by teachers and the impact on the pupils' performance and learning.

As each of the above objectives were discussed with ECZ, it became evident that ECZ should reconsider its overall testing program, from a policy perspective, rather than address or change each of the individual objectives. Therefore, this report will not address each of the above objectives, but instead it will describe the key observations about the overall BCT and IRI testing programs, accompanied by alternative steps that ECZ may wish to pursue. Although the trip focused on grades 2, 3, and 4, ECZ will likely want to revisit its testing programs at grades 7, 9 and 12, especially if they intend to follow similar procedures that are currently being implemented at grades 2, 3, and 4.

Observations About the Current Implementation of the Testing Program

The first phase of the consultancy addressed grades 2, 3, and 4, but grade 4 is currently the only assessment administered within the BCT program, so most of the comments apply to the grade 4 BCT. However, the comments can be generalized to the other grade levels (grades 7, 9 and 12) within the BCT program and to the IRI program. Because documentation about testing

procedures specific to the grade 4 BCT is limited to the *Teacher's Guide*, the needs assessment of the current BCT program was based mostly on meetings and conversations with ECZ. There are several reports written about the use and administration of the tests for the IRI program, but the reports are of limited use in conducting the needs assessment because they do not describe the processes that ECZ uses within its testing program. Therefore, it is not possible to evaluate the impact of the IRI program at this time. Technical manuals and procedural documents should be developed for every key step in the test development process, such as test blueprints, item development and review procedures, forms assembly, test administration, psychometric analyses, and test use and interpretation. Doing so will allow the ECZ to document its procedures for their reference, in addition to ensuring that the procedures are standardizing across grades 4, 7, 9, and 12.

The first step in understanding and evaluating a testing program is to learn the purpose of the tests. The purpose for the Grade 4 Basic Competence Test is to empower teachers with the skills and instruments of assessment so that the progress in pupils' performance on the essential competencies in literacy and numeracy can be regularly monitored and enhanced (handout, *2nd Conference for the Association for Commonwealth Examinations and Accreditation Boards (ACEDAB) Conference, March 2002*). Identifying the students who require remediation in literacy and numeracy is an ancillary component for administering the BCT.

Item Development Process

One of the most important components for a valid testing program is the alignment of a test's purposes with what a test is supposed to measure. The absence of this alignment results in an invalid test, which means that the test is not useful for the intended purposes of the testing program. Discussions with ECZ and a review of the available documents — *Syllabus* (content to

be taught at each grade level) in mathematics and literacy, test blueprints, and sample test booklets — suggest a good alignment among the content, the test blueprints, and the tests in both numeracy and literacy. For example, the grade 2 test blueprint indicates that two items on every test must measure the students' ability to divide by 2, 3, 4, 5 and 10. A review of the sample test booklets in grade 2 shows that there are two division items on the test that require the students to divide by 2, 3, 4, 5 or 10. Because the items measure division, as defined by the *Syllabus*, they are valid measures of divisions. Further, the items are presented in a clear way, which should facilitate the students' ability to answer them (i.e., minimize the amount of nonrandom error that is introduced in the students' test scores).

Teachers attend an item-writing workshop, facilitated by ECZ, and learn how to write items using the *Syllabus* and the test blueprint. They then return home and write items. The teachers reconvene at another workshop where they refine their items. ECZ also facilitates this workshop. During this second workshop, the teachers and ECZ review the items for alignment to the *Syllabus* and test blueprint. ECZ and staff from the Content Development Center make the final decision about the adequacy of items. Under ECZ's guidance, the teachers then assemble test booklets, using the items from the workshop and the test blueprints. The ECZ reviews the test booklets for accuracy.

Currently, ECZ reviews items and assembles test booklets based only on content considerations and expert judgment, and not on technical properties, such as item and test booklet mean difficulty or discrimination (e.g., biserial correlations) indices. Best practices suggest that both the item and test review processes include empirical analyses and expert judges to review items and tests. Therefore, ECZ should consider revising its current practice of not using empirical evidence when reviewing items and assembling test booklets. ECZ also should

consider developing an analysis plan for use by its in-house statisticians, to ensure that ECZ obtains the types of analyses that are necessary to run its testing program. An analysis plan is a common component of a testing program because it helps standardize the types of statistics and procedures that are used to analyze item and test score information. Currently, the in-house statisticians run the analyses independent of ECZ's input. Some statistics (e.g., biserial correlations and differential item statistics) are absent from the current analysis, which implies that ECZ is not getting the necessary statistical information to make evidence-based decisions. An analysis plan will provide ECZ better control and management of its analysis work. Samples of these types of plans are available to guide ECZ's development of its plan.

Test Administration and Score Interpretation

Standardized test administration procedures accomplish many things, but perhaps the most important is that they contribute to a test's reliability and validity. ECZ's current practice of administering tests is not standardized, for either the BCT or the IRI program, therefore ECZ will want to consider revisiting its current test administration process. The first limitation arises from ECZ providing districts with a copy of a sample test booklet (developed during the item-writing workshop mentioned earlier) and giving every teacher the option of administering the sample test booklet to her students or developing her own items and test booklet using *Teacher's Guide*, the *Syllabus,* and the test blueprints in literacy and numeracy. ECZ can immediately improve this limitation by providing the districts with one version of the literacy and numercy tests, which are mass printed and distributed across Zambia.

The second limitation involves the test instructions and supporting materials. Best practices suggest that, in addition to the test instruction being clearly written, other materials such as sample items and criteria for scoring a test, should be included in standard test

administration instructions. ECZ has developed and provided the teachers with test instructions and scoring instructions, but there is no process to monitor whether the instructions are being appropriately and consistently used by the teachers, or regularly enforced by ECZ. Further, the scoring instructions do not tell teachers how to score items that students skipped or how to score items when students provide two answers for one question. Both of these limitations give the test users too much flexibility in determining how to administer and score the tests and therefore introduces statistical variance and error in the test results that limits their usefulness and meaningfulness. However, ECZ can control and improve its current practices by implementing standardized test administration processes for all of the test administrators and test takers.

Without making adjusts to the current administration system to include, at a minimum, the best practices mentioned above, the usefulness and meaningfulness of the results from the current BCT program are weakened. The IRI program is also impacted by the current administration and scoring procedures because the variance in the process is likely masking any true effects the IRI program is having on furthering students' learning.

ECZ offered practical explanations for not standardizing the current test administration process. For instance, ECZ explained that it might not be able to meet a predetermined test delivery date if the test booklets have to be printed and shipped to all (or a representative sample) schools or districts in Zambia. ECZ also explained that the testing office may become defunct in the future, and therefore it wants to build teachers' capacity to develop their own tests (for continuous classroom assessment) should this become a reality. Both of these situations are daunting, practical hurdles for any testing program to overcome.

However, ECZ may prefer to consider policy issues at this point in the testing program, to ensure that the BCT and IRI programs are meeting there purposes and goals. For instance,

other components of an educational system also can be affected when tests are administered using the current ECZ plan. First, because every teacher has been given the option to develop and administer her own competency tests, ECZ and the Ministry of Education are not able to reliably know where to focus professional development activities and funding. Second, the current grade 4 BCT administration procedures make it difficult to accurately gauge Zambian students' performance on the competencies measured on BCT and on the IRI grades 2 and 3 assessments. Third, allowing grade 4 teachers to write their own items and assemble their own test booklets leaves open the possibility of invalid tests being used to assess students' competencies — based on the national *Syllabus* — in their final year of primary school. Each of these components is important in an educational system.

Another area within the testing program affected by the current administration procedure involves the use and interpretation of the test scores. Under the current system, teachers can assess their students' performance on one test administration (e.g., the test administered at the end of the second term), but they cannot compare students' test results with previous test results to look for improvements or declines in student competencies or monitor the impact of instructional strategies. The current system could be improved to allow for these comparisons. For instance, if ECZ administered one version of the literacy and numeracy test booklets at a specified time every term, then all of the teachers within a school or district could reliably assess not only their students' performance but also the instructional strategies and curricular materials implemented across their classrooms and at the school or district level. These types of comparison are possible when all of the teachers use test results from a common test. Further, curricular discussions at the classroom level can be determined under this proposed plan but they

can be made more reliably if examined from the school or district level, because of the larger number of student test scores aggregated (e.g., averaged) at the school or district level.

The above practical suggestion should be combined with a plan to statistically equate the test results from one administration to another. Statistical equating is too technical for this report but, generally, it is a procedure that aligns the students' test scores to the same standard and therefore allows teachers to assess their students' performance over time. The equating model selected is dependant on the test design and test administration procedures. Thus, a discussion about equating will be more fruitful when AIR is informed about ECZ's plan to revise its current testing procedures. Also, AIR anticipates that technical considerations are likely secondary to ECZ at this stage in the testing program. The ECZ is more likely interested in ensuring that the testing program is aligned with Zambia's educational policies and intended purposes. Nonetheless, developing analysis plans to be used by ECZ's statisticians and other technical documents are imperative records to have and maintain in a testing program.

ECZ may wish to consider altering its use of performance-level categories that are being used to define student performance on the BCTs. ECZ has developed a method for teachers to assess their students' performance by giving them performance-level categories and written performance descriptors. This is a typical way of reporting test scores to students and teachers in national testing programs. However, ECZ's current method of calculating the students' test scores and assigning them to a performance level will not likely serve the purposes of the BCT or IRI programs for the following reason.

The current system assumes a normal distribution of students' scores, which is a symmetrical distribution that provides a model of relative frequency distributions. The total area underneath a curve equals 100 percent. When the area is divided into parts, the percentages of

the scores falling under the curve can be determined. There are six parts under a normal curve that are often regrouped in three ranges. About 68 percent of the scores fall within one range (between one standard deviation above and below the mean). About 95 percent of the scores fall within the second range (between the mean and two standard deviations above and below the mean). About 99 percent of the scores fall within the third range (beyond two standard deviations above and below the mean). Using these theoretical ranges, ECZ identified five levels of performance. The ranges for the grade 4 BCT for both literacy and numeracy are:

Level	Constant Standard Deviation from Normal Curve	Test Score Range
1	-1.96	0-5
2	-.98	6-10
3	0	11-19
4	.98	20-27
5	1.96	28 above

The levels, test score ranges, and constant standard deviations do not change from a test administration to another. The test score ranges were determined from one administration of the literacy and numeracy tests to a nonrandom group of about 600 4th grade students. ECZ prepared a written description (performance level descriptor) for each of the five performance level categories to help teachers interpret students' level of competencies within each category. Teachers are directed to use the performance level descriptors to determine which topics in literacy and numeracy their students require remediation (see Annex 5 in *Grade 4 Basic Competence Tests Programme: Teacher's Guide* and section 2.6 in *Learning Achievement at the Grade Two (2) Level in the Interactive Radio Instruction (IRI) Centres*).

Teachers assign their students to a performance level by calculating each student's raw test score (e.g., total number of correct responses) on the numeracy and literacy tests. The teachers then assign each student to one of the five predetermined performance categories.

ECZ may wish to pursue alternate methods of assigning students to performance levels, as the current method is limited both by the test development process (i.e., every teacher can administer her own test) and by the method used to define the range of test scores (i.e., nonrandom sample of 4th grade students). The two limitations will mask improvements and declines in performance over time, for both the BCT and the IRI programs, because performance is always based on mean performance — for every new test and test administration and because teachers are not using a common test to measure student performance on the competencies.

Recall that teachers are assembling tests based only on content considerations, therefore, some teachers may place several difficult (or easy) items on a single test, which will make the test difficult for some students. This situation will result in more students being assigned in the lower performance categories simply because they took a difficult test. Comparatively, more students will fall in the higher performance categories because they were given an easy test. Without adjusting the difficulty of the tests or the range of test scores, several students will be misassigned to performance categories.

The misassignment of students is problematic in any testing situation, but it becomes even more problematic when a test is used to make grade promotion decisions or other kinds of high-stakes decisions. A high-stakes situation arises when a test is used to base a decision that will directly impact a student's educational experience, such as grade promotion, selection into competitive academic programs, or receipt of monetary or nonmonetary awards. The grade 7

BCT is a high-stakes test because it is used to decide whether a student will be promoted to the next grade.

The issue about setting performance levels is too technical for this report, but the current system could be altered by implementing a few alternative procedures mentioned earlier, such as administering one test booklet to every 4^{th} grade student (or 2^{nd} or 3^{rd} grade student for the IRI program) and equating the test scores. The alternative procedures would enhance both the ECZ program and the IRI program. Both programs will obtain more precise information from the test scores, which will, in turn, allow them to better assess the degrees of student performance on the competencies outlined in the *Syllabus.* In particular, IRI will have more precise test information to determine whether the instructional activities broadcast over the radio are having an impact on student learning. The limitations of the current system, which are outlined throughout this report, weaken its ability to judge the effectiveness of the IRI program.

The last area for suggested improvements applies mainly to the BCT program rather than the IRI program as it is currently being implemented. Under the BCT program, teachers can use the test scores to remediatate students. Thus, the BCT has been earmarked to serve as a diagnostic test of student performance on the competencies found in the *Syllabus*.

Diagnostic measures typically consist of several test items within a few topics so that teachers have access to a lot of information within a narrow range of competencies. For example, if a teacher wants to determine whether her students understand addition, she would give students a test with several (e.g., 5) addition items on it, preferably some of the items would be easy and some of them difficult to solve. An examination of the literacy and numeracy test blueprints shows that the tests have too few items on each topic to provide reliable, valid, and therefore meaningful information that teachers can use to plan a remediation strategy. For

instance, the grade 3 numeracy test blueprint has one item for the topic of Addition and Subtraction. If students answer the one addition or subtraction item incorrectly, it does not necessarily indicate that they are incapable of adding or subtracting. Rather, it could likely mean that the one test item did not function properly—that is, the item could be a "bad" item. Further, if teachers use the results from one item to make remediation decisions, they may be using valuable resources to reteach students about topics on which they may not require remediation.

If ECZ would like to continue using the BCT as measures of remediation, the tests should contain more items for every topic than they currently do so that the teachers can obtain the information they need to make informed decisions. This reconceptualization of the BCT would require lengthening tests from 20 items to at least 55 items, because there are eleven topics covered on the tests. It also would require additional testing time and an investigation into whether a sufficient number and diversity in type of items can be written for every topic in both literacy and numeracy.

Possible Solutions and a Way Forward

The needs assessment has identified one area that is using adequate procedures to develop the BCT and the IRI tests (e.g., item development) and several procedures that should be improved and even altered (e.g., forms assembly, test administration, and score interpretation) to meet the purposes and goals of the overall testing program, both for the BCT and the IRI programs. However, before ECZ considers making refinements to the current testing program, AIR suggests that ECZ first conduct an overall policy review of the program, as some components of the current system may not be addressing or furthering the original goals and purposes. Some of the questions that ECZ will likely want to consider include:

- How well are the policies and current testing practices aligned?

- Do some of the current test development, test administration, and score interpretation policies need to be reconsidered?
- Does the testing program need to be better aligned with current policies?
- What is the ECZ's goal and purpose for measuring competencies on the BCTs?

If ECZ determines that the current program is indeed what it originally envisioned, then refinements to the current procedures can be implemented to enhance the current program. For instance, ECZ should include the use of empirical evidence to review items and test forms. It should also consider mass printing and administering one test booklet in literacy and numeracy per an administration to eliminate teachers having to develop their own BCT tests. This option does not have to result in ECZ abandoning its wish to provide teachers' with professional development in the area of developing teacher-made tests. The professional development activities would switch from focusing on the BCTs and monitoring students' attainment of the competencies on a term-by-term basis, to classroom based testing (or continuous based testing) and the monitoring of student performance, perhaps weekly or bimonthly, using teacher-made tests. Regardless of the ultimate decision, AIR would like to suggest a few options for a way forward for ECZ's consideration.

- ECZ and other stakeholders should convene a one -or two-day meeting to revisit the policies and purposes of the testing program. This meeting would allow ECZ to "take a step back" and revisit the initial goals and plans for the BCT program and assess whether or how it is currently satisfying the initial goals and plans.
- ECZ should determine whether the testing needs for the IRI program are being adequately satisfied.
- ECZ should consider having the teacher-training institutes instruct prospective teachers how to develop, administer, and score continuous assessments or have separate in-service workshops, to differentiate between competency based and continuous assessments.

- ECZ should consider its staff not be involved in assisting teachers with continuous assessments so that they can focus and dedicate its resources on competency based testing.
- If ECZ determines that the current system is meeting its goals, then ECZ could work to refine and improve the current BCT system, based on outcomes of the needs assessment. For example, ECZ staff specifically requested assistance in establishing item banking and tracking procedures, establishing performance level descriptions, considering policy issues and how they relate to high stakes testing (e.g., grade 7 BCT), ensuring that best practices are used throughout the entire process of developing the grade 7, high-stakes tests, using test scores to provide useful and meaningful information to teachers, reviewing the alignment between the revised curriculum standards and the test blueprints, establishing guidelines for developing analysis plans, and identifying the documentation necessary for a national testing program.

Next Steps

ECZ may wish to pursue next steps in the process of developing competence based tests, for both the BCT and the IRI programs. AIR would like ECZ to consider the following next steps, which would occur in the near future:

- Conduct a review of the tests and the procedures used in development, administration, and scoring in grades 5 and above. It would be similar to the review conducted for grades 4 and below. This review would require a visit to Zambia to conduct an in-depth review of the tests in grades 5 and above and identify and arrange meetings with key developers and users of the tests.
- Develop training materials based on the visits and return to Zambia for a second visit. The second visit would include a workshop in examining best practices in assessment and a meeting to decide how to proceed with the recommendations from the trip reports.
- Meet with the developers and users of the IRI program to discuss specific strategies for assessing the impact of the program and possible research designs to precisely measure impact. Determine ways to move forward with the IRI program, for instance should it continue to use existing tests at the lower grade levels or is there a need to develop

alternative tests. Administer tests to a sample of students and analyze results. Convene a one- or two-day workshop with attendees yet to be determined to present results about the impact of program.

- Develop a schedule of testing activities to schedule future workshops, for both BCT and IRI needs.

References

Basic Education Mathematics Syllabus (Grades 1-9). Examinations Council of Zambia, prepared by the Curriculum Development Centre, Lusaka, Zambia.

Grade 4 Basic Competence Tests Programme: Teacher's Guide, (2002). Ministry of Education,Lusaka, Zambia.

Learning Achievement at the Grade Two (2) Level in the Interactive Radio Instruction (IRI) Centres. Examinations Council of Zambia on behalf of the Education Broadcasting Services, Lusaka, Zambia.

Mutanekelwa, N. N. and Kapambwe, W. M. (March 2002). 2nd Conference for the Association for Commonwealth Examinations and Accreditation Boards (ACEDAB) Conference. Handout from the ACEAB Conference.

Zambia Basic Education Course: English Syllabus (Grades 1-7). Examinations Council of Zambia, prepared by the Curriculum Development Centre, Lusaka, Zambia.