

Exam Fervor and Fever

*Case Studies of the Influence of Primary Leaving Examinations
on Uganda Classrooms, Teachers, and Pupils*

Volume One

Conrad W. Snyder, Jr.

1997

Contents

Acknowledgments	iii
Acronyms	iv
Executive Summary	v
Introduction	1
Influence of Exams on Classroom Dynamics	2
Primary Leaving Examinations in Uganda	12
<i>Teaching and Learning</i>	13
<i>Case Studies of Schools</i>	15
<i>Influence on Pupils?</i>	17
<i>Influence on Teachers?</i>	20
<i>Some Quantitative Details</i>	23
<i>Influence on Classrooms from the Teachers' Perspectives?</i>	25
<i>Technical Limitations to the Use of Examinations as a Lever for Reform</i>	28
Strange Loops, Complexity, and Institutional Credibility	33
Conclusions	39

Tables

<i>Table 1. Characteristics of Test Format</i>	42
<i>Table 2. Tests in Recent Reforms in the United States</i>	45
<i>Table 3. Rubrics for Observation Variables</i>	48
<i>Table 4. Pupils' Position on Abolishing the PLE from Their Essays in the 1995 PLE</i>	52
<i>Table 5. Means and Standard Deviations for Classroom Observation Items for 31 Classrooms in Uganda</i>	53
<i>Table 6. PLE Characteristics Linked to Teachers and Pupils in Teachers' Interview Statements</i>	55
<i>Table 7. Summary of the Behavioral, Institutional, and Cultural Dimensions for the Uganda National Examinations Board Examinations Reform</i>	58

Acknowledgments

USAID's Bureau for Africa initiated and funded the study of the Uganda examination reform through the Advancing Basic Education and Literacy Project. The study was carried out under the auspices of the Academy for Educational Development. David Chapman and John Hatch managed the project from AED, with very capable support from Monica Fox. Our special thanks to Monica for her many extra efforts in communication and coordination.

The design and technical aspects of the project were coordinated by Wes Snyder and George Johanson, with considerable assistance after the field work from Bonnie Prince. The field team was headed by Cooper Odaet. His efforts included all field coordination, access permissions, logistics, and management of the data collection. Our successes reflect his extraordinary concern for Ugandan education in general. Interviews with administrators were handled by Lazarus Jaji, and classroom observations, by Meriwether Beatty. Bonnie Prince helped in the editing of interview protocols, carrying out interviews with teachers and handling the massive qualitative data analysis of the interview information. She continues to analyze the rich qualitative data, even as this report is being written. George Johanson carried out many of the quantitative analyses. Wes Snyder wrote the final report, Volume One. Bonnie Prince collated pupil essays and teacher interviews in Volume Two.

We are most fortunate that the Uganda National Examinations Board opened its files to give us the pupil essays on abolishing the Primary Leaving Examinations (PLE) from the 1995 PLE scripts, and teachers, headteachers, pupils, and parents at the selected schools gave their time and ideas to thoughts about the PLE, its role and its impact. The interviewees were thoughtful and insightful. They shared our commitment to better understand national examinations. This study is based on the views and ideas of those in selected schools of Uganda. We hasten to add that this study is not an evaluation of the Ugandan education system. Our intent was to look at particular case studies of schools possibly affected by the creative examination changes instituted by UNEB over the last five years.

We dedicate this report to UNEB. Their courageous and well-intentioned examination reform has provided a guide to us all. There is still much to learn from them.

Acronyms

ABEL2	Advancing Basic Education and Literacy Project, Phase 2
AED	Academy for Educational Development
ASAP	Arizona Student Assessment Program
CPE	Certificate of Primary Education
CRESST	National Center for Research and Evaluation Standards
ECAPE	Exploratory Committee on the Assessment of Progress in Education
EDI	examination-driven instruction
IEQ	Improving Educational Quality Project
MDI	measurement-driven instruction
NAEP	National Assessment of Education Progress
OLB	“one less bump”
PLE	Primary Leaving Examinations
UNEB	Uganda National Examinations Board
USAID	U.S. Agency for International Development

Executive Summary

In many education systems, examinations define success and failure for individuals, act as gatekeepers to future opportunities, and provide credibility to the systems that engage them. Uganda's Primary Leaving Examinations (PLE) exemplify these features: they signal the end of seven years of education, document the skills and information the pupils can demonstrate on this occasion, presumably as a result of schooling, and present a high-stakes hurdle for continuing formal education. The PLE have survived war, poverty, and isolation in a country that has only recently found political stability. Given the widely recognized influence of examinations and the obvious importance of end-of-cycle decisions, the question of the usefulness or damage of that influence is an important one for those interested in an effective education system, as well as for those who wish to find a way to influence that system. Examinations appear to be essential features of credible systems and a potentially powerful influence on anything that will happen in the instructional program. Given the general influence of examinations, how do they affect the children and adults who are involved in them? Can the specific nature of the examinations influence the way education is delivered? Are there simple strategies that can be used to manipulate the examination approach to have focused impact on the classroom?

The answers to these questions, as suggested by the study of 16 schools, 60 teachers, 15 headteachers, 31 classrooms, and many community members and parents in Uganda, are that examinations have substantial effects on the individuals involved. In fact, the effect may be greater in emerging education systems than we ever realized. Second, examinations do influence the way education is delivered. Again, the influence is not necessarily as we expected. And third, the impact of examinations is not simple. Changes in content and approach to the examinations, coupled with information disseminated to teachers and some inservice attention to the changes, do not offset the formalities of schooling, which include the rituals of "exams." Examinations are more than their parts. National examinations are poignant events in the lives of schooling participants, even before they occur. The social meaning of examinations goes beyond the determination of which children qualify for continuation, which is important enough. Examinations appear to be linked to universal notions of schooling that transcend local conditions and actions. In Uganda, the Ugandan National Examinations Board (UNEB) attempted to alter the content and process of the examination questions in order to "direct" the kinds of changes hoped for in classrooms. But the social significance of examinations exceeds the technical qualities of their form so that simplistic strategies to influence classroom dynamics will not work directly. By virtue of their international identification, examinations accredit systems and individuals. Their influence is dramatic, but not easily controlled or even understood.

The implications of this work are that, for contexts similar to Uganda, interventions like improved school management, better instructional strategies, new syllabi, installation of continuous assessment schemes, and technical adjustments to exam content or process are sideshows to educational reform. Examinations define the nature of schooling, not solely in the technical sense as we logically expected, but in a social sense that is even more persuasive and influential than a technical adjustment. All aspects of the system remain robust and resistant to

long-term reform, and the examinations lie at the heart of this robustness. Alternative education assessment strategies address technical issues, but fail to deal effectively with institutional and credibility concerns. Examinations are part of what schooling is, and their absence would not be tolerated. But can examinations be used to reform a system?

We started out to identify the effectiveness of UNEB technical interventions. They can be summarized as cognitively strong and dynamically weak, but we found the dramatic “reach” of examinations in the lives of schooling participants as the dominant feature of this high-stakes set of events. Examinations are associated with mental/emotional, physical, social and family, learning and teaching, and systemic dysfunctions. While the children and teachers have different meanings for “exam fever,” fever emerges as the umbrella metaphor for the whole range of dysfunctions that embrace examinations. Still, particularly for teachers, examinations are the frame for thinking about the schooling experience. Examinations determine the credibility of the system, and are credited with defining national aspirations and accomplishments. And a majority of teachers support the use of examinations, despite their side-effects. Reference to the quality or capacity of the education system depends ultimately on the perception of the examination system and performance on its terms.

Despite criticisms and problems, examinations are the standard setters for schooling. An examination is an awesome lever for reform, but not necessarily along the technical lines that a rational analysis would have expected and perhaps not as linear or predictable as we might have hoped. In terms of House’s application of transaction-cost economics to educational reform (House, 1996), examinations as a reform device fall short of required technical attributes. As we found in Uganda, teachers are subject to *bounded rationality*—teachers cannot necessarily adjust to examination changes; even if they understand the examination requirements at a cognitive level or an emotional level, they may be able to do very little in making the necessary changes in the classroom to affect performance improvements. Also, examinations are already subject to *opportunism*, particularly in Uganda. Alternative assessment strategies would have no chance of surviving the level of individual motivations to look good and influence local results. Despite the security involved in the PLE, the common frames for essays divulge a great deal of collusion and cooperation for local advantage. Examinations do allow *asset specificity*, thus meeting only one of the required attributes for probable effectiveness as a reform device, according to House. This Ugandan study verifies what House would have predicted. But these are only the technical dimensions of reform. As House points out, there are political and cultural aspects for reform that must be taken into consideration. We started out looking at the technical aspects of the PLE, but ended up recognizing the awesome power of the PLE in political and cultural terms.¹ Examinations alone will not be sufficient to alter the incredibly robust education systems that require reform. But exams and assessments are essential features of any reform strategy, particularly in emerging education systems, where examinations determine individual futures and define the very notion of schooling.

¹House, E. (1996). A framework for appraising educational reforms, *Educational Researcher*, 25 (7), 6-14. House advocates self assessment and “relatively autonomous and unregulated” schools. Schooling in emerging systems is more centralized and less likely to achieve needed credibility to operate outside the national system. Schools remain decoupled from central authorities but “voluntarily” submit to international standards and examinations in order to certify their pupils and programs.

F

One of the problems in understanding education reform is that institutional success is not always congruent with technical progress. Examinations legitimate the educational enterprise in Uganda, and Ugandan schools and teachers have survived enormous stresses and incursions. Uganda has a long tradition of education that is often reflected in individual successes from those succeeding in “high standards” examinations. But the examinations remain remote visions to the teachers of Uganda. Individual advantages tend to go to those who achieve well throughout their educational careers and may have less to do with formal instructional programming. At the higher cognitive levels, there is no particular linkage between what an examination asks and what a teacher does. As the testing literature documents, specific changes in an examination, especially in recall items, can emerge in classroom coverage. But complex cognitive processes demand different cultural and dynamic features in classrooms that are unlikely to emerge spontaneously. Examinations are essential for reform, but not sufficient. They mark the legitimization of the system as well as its technical competence.

Advocates for alternative educational assessment recognize these deficiencies in high-stakes, single-occasion, narrowly derived measures of performance and understanding. But they fail to fully recognize the role of national examinations in the survival and “success” of emerging central education systems. Examinations are part of the ritual of schooling. Potentially dramatic individual dysfunctions are tolerated because of the mobilizing and accrediting aspects of examinations to substantial parts of the internal polity and the universal educational community. Organizational systems are highly dependent on their political and cultural environments and on the support and legitimacy that they are accorded in those environments. Examinations “guarantee” that legitimacy, regardless of their connection to actual technical aspects of the system. Emerging systems, in particular, are likely to exhibit considerable fragmentation and inconsistency among their organizational components designed to implement the educational mission. This “loose coupling” can be an advantage where many elements are deficient and resources for many activities inadequate. National examinations “permit” the continuation of many different and uncoordinated efforts to continue without system disintegration. The Ugandan education system has survived nearly unimaginable difficulties over the last several decades and examinations have defined the system in the absence of many, usually considered important, features and resources. Rationalistic reformers cannot afford to overlook the ritual and power of examinations.

Our conclusion is that national examinations are complicated components of a complex system. The UNEB reform has captured the attention of teachers, headteachers, pupils, and parents, but the effects of a change in perspective to a *thinking* classroom are not yet apparent. Although national examinations provide a lever for central policy control over instruction in the classroom, their use in reform is not as simple as it might first appear. Exams are important aspects of reform, but not sufficient. Education is not altered easily or predictably by simple solutions.

Introduction

First came the economists, then came the invention of the wheel—efficient transportation, added value for carrying goods, and new markets for carts. Of course it took a few hundred thousand years to work out the bugs, slowed perhaps so that the market would prosper for a long time. The “wheels” were square, and people complained about the bumpiness of the ride.

One fine entrepreneur/inventor displayed her newest invention. “Here,” she announced proudly to potential distributors, “is my new, improved wheel. I call it the ‘rolleasy.’ Isn’t it a beauty?! Ten skins (manufactured of course) I’m going to charge. To the first of you, my friends, a real bargain at six skins each, provided you buy a set and spare for each of the carts you sell.”

The distributors looked at the new wheel with some puzzlement, as frequently happens in economic contexts. Eventually, one said, “But it’s triangular.”

“Of course,” replied the profit-oriented inventor.

“How can that be an improvement?”

“Don’t you see?” expostulated the inventor. “One less bump!”²

From the outside, the “one less bump” (OLB)³ triangle appears to be an improvement over the four bumps associated with the square. But if you’re inside, the bumps come slower but more dramatically.⁴ You’d hardly characterize fewer and larger bumps as an improvement. In fact, if you’re modifying a square for less apparent ‘bumps,’ you’d really want to increase the number of bumps, gradually increasing the number of sides from that of a square and approaching the form of a circle, which we know now to be a better solution (and when attached to a vehicle, skins our bank account). So you wouldn’t expect the OLB to catch on. But it appeals, at first, to our logic—one less bump, and therefore, implicitly, a smoother ride. And advertising would surely impress this point. We’d probably buy it for awhile.

Education reform is an enterprise always looking for OLBs. We are reluctant to admit to the complexity of education, particularly its social and political dimensions. So, it’s not surprising that new reforms are frequently touted as OLBs. Reformers have looked everywhere for “rolleasy,” apparent fixes that imply a smoother ride to promised outcomes. But since each solution simplifies the complexly organized system of problems to some particular single

²Based on a story told by J. Cohen and I. Stewart in *The Collapse of Chaos*. New York: Viking, 1994, p. 366.

³Not to be confused with OLS, ordinary least squares. OLS has all the bumps!

⁴Of course if the triangle is an equilateral triangle such that a side is equal to that of the square, then there would be no noticeable difference. Either way, the ride continues to be bumpy, despite the OLB.

dimension of the set, it is an OLB. It appears smooth, but in fact the myopic solution frequently exacerbates the situation by failing to consider the interdependency, multidimensionality, and insolubility of educational problems. The one thing we've learned (well, not many of us) is that education must be considered complex.⁵

One of the more recent OLBs is measurement-driven instruction (MDI). The approach is as described—exams or tests directly drive what happens in the classroom. The approach presupposes a simple set of independent elements for instruction: teacher, pupil, curriculum, materials, and exams. We have unilaterally educated teachers with some effect; we have selected pupils with controversial success; we've improved, enhanced, and expanded curriculum with dubious results and greater responsibilities for modern schooling; and now exams—if tweaked, they will, it is claimed, because of their importance (high stakes), influence what happens to what the teacher teaches, what the pupil focuses on to learn, and the emphases in the curriculum. For those who view education as complex, MDI and its cousins are OLBs. Exams, particularly high stakes exams, have effects, but the story just isn't as simple as we might like, and they cannot be treated in isolation of other aspects of the education system and even beyond. This doesn't mean that exams are not a good lever for reform. As above, we have to add more dimensions for a more effective “ride.”

Influence of Exams on Classroom Dynamics

Much of what constitutes research on classrooms has focused on isolated classroom behaviors of teachers or pupils. The results vary from context to context, and the specifics of “best practices” seem to depend upon local cultural characteristics and expectations, as well as subject matter traditions, as much as anything else. Furthermore, schooling is a political and social process with ambiguous, or at least debatable, technical features. The causal linkages between the technical aspects of teaching/learning and outcomes are weak, particularly at the complex cognitive levels. Many practices work, for a variety of reasons, in some contexts, and yet may not work under other circumstances. Success is dependent upon how the practices are carried out, the context within which they occur, and many other qualities of person and facilities that are difficult to delimit in any precise formula for success. The dynamic aspects of the learning events seem more important than any formula that emphasizes specific elements might capture. There are many arrangements under which classrooms can be a success, and there are many kinds of successes.

In this paper, the classroom is seen as an affective ecology within which contiguous interactions evolve from historical, social, and contextual actions. This is in contrast to the usual

⁵In dealing with complexly organized system of problems, we need maximum information. This means constant research, monitoring, and evaluation. This is not always valued in solution oriented projects. We continue to seek the OLB in the development context. Because the solutions have to be easy to understand to explain within the political arenas of development, there is a natural bias away from complexity.

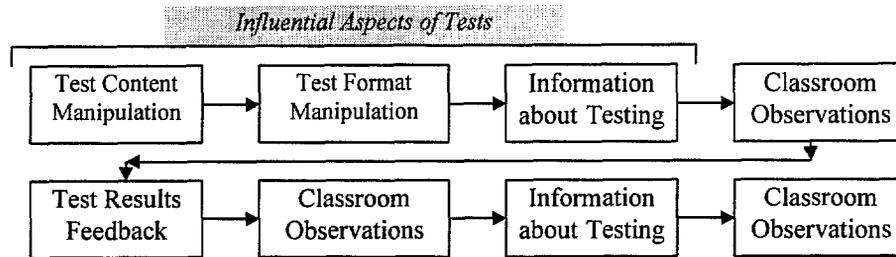
depiction of what to look for in classrooms to explain outcomes.⁶ The key components of classroom instruction are not distinct, static, “independent variables” but rather dynamic variables that follow the ebb and flow of classroom activities and atmosphere. Dynamic variables are conceptual categories for complex arrangements and mixtures of events and reactions that are repeatedly changing and interacting with other variables. Teachers and pupils present a complex social picture, even for a single lesson. Since micro-behavioral analyses have not been particularly illuminating, classrooms may be better characterized in terms of perceptual-action cycles that continually evolve, patterns emerging and then dissipating into the backcloth of activity, as new patterns present themselves. Single lessons are samples of the general flow that characterizes a particular teacher-class combination, and the variables attended to are macro collectives for larger and complex trends of activity.

A key consideration in the flow of classroom activities is the scope and direction taken by the teacher. At one time this was thought to be related solely to the instructional and curricular decisions made both in planning and carrying out the classroom program. Now we realize that many factors feedback into the classroom process to affect the dynamics of classroom interactions. One apparently powerful influence is the high-stakes testing that some national systems engage in for selection and evaluation. Although testing has been an important part of modern instructional programs since the late 1960s, the initial claim was that participation in a course affected the scores on a test, and there was little concern for the effect of a test on a course. Feedback on testing did inform those interested about the effectiveness of the course, but little emphasis was given to the influence of testing itself on the choice of material or pedagogy. Over time, testing has received more attention as an important influence on what happens in a classroom. The concern was expressed in terms of the limited scope of the achievement testing, usually framed within a single-answer, multiple-choice format. The focus on classroom influences drew attention to the limits of testing frames on instructional choices and the overall influence of testing on future versions of a course.

Measurement-driven instruction (MDI) emerged from this debate and emphasized the positive potential of testing to directly shape teaching and learning. Criterion-referenced tests articulated the behavioral approach to scope and instructionally defensible targets, and later, performance tasks broadened that view to include more cognitive and expressive concerns. The original argument becomes better formed: The course influences performances in a test, where the test measures those aspects related to what goes on in the course, and the nature of the test, in turn, influences what goes on in future versions of the course. Testing influences include test content, test format, and information about test content, format, and problem areas.

⁶Classroom observations for explanation are of course different than observations designed for comprehensive diagnostics. In this case, we are looking at a small set of macro process variables that are implicated in classroom effectiveness.

Figure 1. Long-Term Design for Measurement-Driven Instruction (MDI)



The nature of a test includes its *item format* (true-false, multiple-choice, essay or short-answer, project or performance-based, dialogue-oral), *content* (topics and subject areas), *level of knowledge called for* (retention, understanding, or utilization), *complexity* (number of content areas and their interrelationships), *difficulty* (easy or hard), *discrimination* (in terms of some cutoff of accomplishment), *referential source* (criterion-referenced or normative referenced), *purpose* (pupil performance, curriculum assessment, teacher competence, school effectiveness, etc.), and *type of items* (achievement or aptitude). See Table 1 for some other characteristics that relate to item format. Because of limited time and examinee energy, tests generally comprise samples of their domain of potential items. Tests are, therefore, always limited in some way, and the idea of a single test to characterize the accomplishments of a difficult set of processes or decisions is not educationally sound.

There is considerable debate about whether schooling should focus on the development of abilities that may be independent of content (aptitude) or embedded within content (achievement). A similar debate surrounds the issue of context in testing: Generalization from artificial schooling conditions (formal tests) or tasks drawn from conditions more similar to the actual context to which generalization is intended (authentic, performance tasks). The validity of the examinations, in national high-stakes contexts, is not always given the attention we might expect. These debates muddy the reform waters even more in terms of what we might expect from the influence of different kinds of tests. The effects are not always what we expect, and they are clearly more complex than we anticipated.

As conveyed in Table 1, it is most easy to link tests with instruction at the level of retention. However, the grander intent of MDI is to explicitly tie testing to higher level cognition and accordingly, influence what happens in the classroom. Linkage is not easy. Teachers may very well see the need but be unable to change their methodologies in the classroom because they lack the requisite professional skills or subject area repertoires. This occurred in Botswana when the history examination of the Junior Certificate Examinations was changed to reflect concept development and analysis. The resulting score distribution was highly positively skewed, with most scores near chance. These results caused considerable political concern because the exams were needed as part of the selection process for the higher forms. Careful staff development and

information dissemination had been carried out, according to the legend.⁷ The exam was inappropriate for selection. Teachers had not adapted to the anticipated higher-level content of the exam because they did not know how to alter their classroom approaches. Although the staff development exercises informed them about the changes, little emphasis was given to the implications for classrooms and their pedagogy and no follow-up in the classrooms was attempted to support the reform. Unfortunately, the attempt was dropped immediately. No further effort was made to examine higher cognitive processes. Instead of making corrections, the politics of selection dominated the decision. Moving back to the comfort of what was easily possible was more attractive than tackling the difficulties of pedagogical change. The teachers presumably recognized the changes in the exams but never intended to change their teaching or classroom dynamics. They merely hoped for the best.

In a more carefully studied example,⁸ the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) carried out three-year, multiple case studies on elementary school reactions to the Arizona Student Assessment Program (ASAP). The ASAP mandated changes in classroom instructional practices in response to new tests that were designed to assess these changes. Case studies yielded information on perceived meanings of the test mandate and the evolution of these meanings over time. The transformation was accomplished in only one school (and this school had already adopted the practices that were subsequently mandated). As the researchers said: “Local interpretations and organizational norms intervened to color, distort, delay, enhance, or thwart the intentions of the policy and the policy-shaping community.” Relationships between examinations, classroom dynamics, and learning are very complex.

The Somerset studies in Kenya⁹ laid the groundwork for the enthusiasm of testing as a lever for reform in the classroom. Tests have always been closely tied to curriculum as measures of accomplishment on the content and processes intended for instructional development, but the new hope has linked testing and instruction in broad reciprocal ways that implied that either element could be tweaked for effect. This enthusiasm was codified in the measurement-directed instruction (MDI) movement, promoted by Popham and others.¹⁰ Borne within the behaviorist tradition, MDI advocates the clarification of instructional targets and the development of criterion-referenced tests to assess the accomplishments in meeting those targets. MDI “occurs when a high-stakes test of educational achievement, because of the important contingencies

⁷Personal communication with Bryan Axtell, long-time consultant to the testing division of the Ministry of Education in Botswana.

⁸Smith, M.L., Noble, A.J., & Cabay, M., Heinecke, W., Junker, M. S., & Saffron, Y. (1994). *What happens when the test mandate changes? Results of a multiple case study*. CSE Technical Report No. 380. Los Angeles, CA: UCLA Center for the Study of Evaluation.

⁹Somerset, A. (1983). *Examinations reform: The Kenya experience*. Washington, D.C.: World Bank.

¹⁰Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, MDI Pro/Con, 679-682. Popham, W.J. (1993). Measurement-driven instruction as a “quick-fix” reform strategy. *Measurement and Evaluation in Counseling and Development*, 26, 31-34.

associated with the students' performance, influences the instructional program that prepares students for the test" (1987, p. 680). The advantage of MDI is that "[i]f properly conceived and implemented, measurement-driven instruction currently constitutes the most cost-effective way of improving the quality of public education in the United States" (1987, p. 679; i.e. a lower-cost OLB). The influence of tests on instruction results from the correspondence of targets between the tests and the instructional program. Tests are a powerful "curricular magnet." Popham offered some advice on the responsible use of MDI:

- ◆ MDI must be shown to be effective in each case.
- ◆ MDI will not lead to curricular reductionism, which emphasizes only the most trivial and least defensible goals, if the tests are aimed at genuinely higher-order skills (e.g., California and Pennsylvania use "honors" tests).
- ◆ MDI will not lead to curriculum stagnation, in which the tests lock educators into an unalterable pursuit of fixed content, if the programs are continually reviewed and developed (e.g., programs in New Jersey and Connecticut have updated their programs to include more demanding skills).
- ◆ MDI does not constrain teacher creativity but it may lead to more focused use of limited resources.
- ◆ MDI is most effective if criterion-referenced tests are used, content is defensible, a manageable number of targets are assessed (between 5-10), the tests help illuminate the curriculum and its instructional intent, teachers use the information to assess progress efficiently, and the tests help to clarify the instructional program.

Popham's early advocacy for MDI was tempered by the large number of negative experiences accumulated over time, when many state-level and local-level mandates and implementation of MDI yielded sour effects. By 1993, Popham was advocating proactive communication by those who are experts in MDI with the influential educational policymakers. His summary provides the basic motivation for the continued consideration of MDI (1993, pp. 25-26).

...all quick fixes are not created equal. Measurement-driven instruction is an educational reform strategy that, if properly conceptualized and implemented, can lead to both immediate and long-lasting educational dividends. If not properly conceptualized and implemented, however, measurement-driven instruction is apt to be a "quick-fix" in the most negative sense of that expression.

MDI work in the United States illustrates two points, one expressed by Popham and the other by a prominent critic of MDI, George Madaus.¹¹ First, high-stakes testing can influence instruction. And second, the influence will change over time, both in quantity (and possibly direction) and the nature of the effect. For those who advocate MDI, if the:

¹¹Madaus, G. F. The influence of testing on the curriculum. In L. Tanner (Ed.). *Critical issues in curriculum* (87th yearbook of the National Society for the Study of Education, Part I, 83-121). Chicago, IL: University of Chicago Press.

- ◆ attributes to be tested are well selected and important; and
- ◆ tests are valid in terms of measuring the attributes chosen;

then the

- ◆ goals of instruction are accordingly clear and discernible;
- ◆ teacher and pupil focus their efforts on these targets;
- ◆ standards are explicit and uniform for each pupil;
- ◆ accountability is easier and more objective at each level of the system; and
- ◆ the public has more information on the quality of schooling.

Madaus points out that the power of tests to influence instruction is a perceptual phenomenon—if you believe it does, then it does; the effect is produced by the perception, regardless of the reality of the importance of the linkage. So the effect may be present even if the attributes are not well chosen and the tests are not particularly valid. When they are both, the effect may be temporary. The second point, which Madaus relates to a “social version of Heisenberg’s uncertainty principle,” is that effects decay or change over time because quantitative indicators will eventually distort or corrupt the social processes they are monitoring. For example, a school that attains a low set of scores on a high-stakes test may institute changes that improve the instruction or they may more directly address the way or what the indicator indicates.

To return to the development context, negative backwash effects were the specific reason for Kenya’s reform of the Certificate of Primary Education (CPE) examinations. The concern was that the range of competencies within the CPE were restrictive and limiting instructional targets. The reform began in 1974 and was particularly influential between the years 1975-1977. The CPE group (Somerset joined in 1977) instituted two changes: revision of content and development of an information architecture for test information. The information related to incentives, consisting of comparative performance statistics, and guidance, consisting of analyses of performance. A newsletter provided information on exam content and diagnostics about exam performance, with suggested remedies. Effects of the reform were complex. At first, the comparative advantage of previously high scoring schools seemed to increase, then the trend reversed in later years. The weakest district improved significantly over five years and became one of the strongest performers. Kenya restructured its education system in 1984.

The lessons from these studies for use in the reform of education systems can be summarized as:

- ◆ The sphere of influence in education is ritual and symbolic. Testing will have an effect, like any other intervention, if the teachers and pupils believe that tests are important and that what happens in a classroom is related to test performance.
- ◆ Because the effect is based on specific tests, everyone will focus on attributes of the tests and not on any technical relationship between instruction and assessment. Therefore, any tests will eventually narrow the focus of effect.
- ◆ Test mandates can be a “quick fix” if effectively designed and implemented, but the fix depends ultimately on the support system that can continually revise the nature and content

of the tests to assure relevance, quality, distribution of quality, equity, and reliability (as per Kenya goals).

- ◆ Regardless of the quality of tests, testing will reflect the existing explicit agenda of an education system and ignore or leave out many of the important aspects of education, particularly in its modern form. Portfolio and authentic assessment strategies have been developed in the United States to offset the narrowness of focus. The broadening of emphases increases costs and requires higher quality teachers to implement. The impact of tradeoffs is not well known.
- ◆ Test formats are important, and attempts to measure high-level cognitive skills are complicated and expensive. Test development costs and technical requirements are high. There is no easy path to good testing.

We can put the problem in strategic terms, as if strategic planning were actually a part of education reform. The difficulty for education reform is that much of what we know about the development of strategy is strategy-as-position, where we can determine solutions for particular circumstances (literally particular products in particular markets).¹² In a sense, we're looking for the easy solutions, or technical OLBs. The details of formalizing the practicalities are more easily articulated in strategy-as-position, so we follow that path. Our real problem, as widely recognized even if we don't act on it, is that of how we do things in education, that is, strategy-as-perspective. We go about schooling as if our intention was merely the dissemination of information, or as Paulo Freire has popularized, "banking education." Our real intention is much more complex. As an example from testing:

Figure 2. Strategy as position or perspective.

		<i>Strategy as Perspective</i>	
		Old "Learning Facts"	New "Learning to Think"
<i>Strategy as Position</i>	New	Efficient Formats	Portfolios
	Old	Essay Exams	Tests of higher level cognition

¹²This section attributable to Henry Mintzberg's (1994) *The rise and fall of strategic planning*. NY: The Free Press, 27-29.

The change from expensive and time-consuming essay exams to short-answer or multiple choice formats for exams is a change in position, where we are concerned in each on “how much a child knows.” But when we wish to move beyond that to assess the ability of the child to think, then we need to change the perspective on how we go about schooling. This is essentially the notion of modern reform: to change schooling fundamentally. Changing position within a particular perspective may be much easier than changing perspective. Within reform agendas, we have also clearly favored strategy-as-plan (intended strategy) over strategy-as-pattern (realized strategy, which includes unintended, emergent dimensions of change). As Mintzberg goes on to say (p.29):

Our conclusion is that “strategic planning” cannot be synonymous with strategy formation, which encompasses all of these, and certainly not with effectiveness in that process. The implication of this is that planning may have less to do with strategy making than is often claimed, but also that planners probably have more work to do than they sometimes realize!

In education reform terms, we know something about how to move within *perspective* to change *positions* and we plan projects in considerable detail, including those involved in developing nation contexts. We know very little about how to alter *perspective*, and we’re not only naive about patterns, but frequently explicitly blocked by proposed plans of deliberate strategies to follow the emergent strategies that are revealed in process. This ties to Madaus’ observation that effects can decay or change over time. We cannot merely introduce an intervention and stop there or wait there. The initial conditions first faced by the intervention are altered forever and possibly, dramatically by the intrusion. Thus, strategy is not formulated as much as formed. This bodes poorly for OLB strategies for exam influence. Strategies in education reform must deal with complex contexts that are changing, and with exams, those changes have focused on the exams themselves and not the intents of the exams. The grander vision decays.

Over the last several years, the recognition of the deeper problems of schooling has led to more dramatic reformations involving changes in perspective. Standards and exams have been at the heart of these schools’ reforms. By estimate, about 41 million children in the United States take approximately 124 million tests annually in their schools.¹³ There is now a different view of what pupils know and how to measure their potential and their accomplishments. This is a dramatic shift in what schools must do to be effective. The question is how to harness the power and influence of tests to provide a “common school” experience for all pupils to reach new levels of performance. This “revolution” is evident in many contexts. Table 2 outlines these new programs and specifies the kinds of interventions that were attempted and the reaction to them. They moved to new assessment procedures, totally revamped instructional and administrative agenda, and faced different political rhetoric.

¹³Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco, CA: Jossey-Bass Publishers.

The move to new assessment approaches is not surprising, given the powerful side effects of high-stakes, single national examinations. Longer term, deeper, and more meaningful performance measures better reflect the instructional objectives than the minimal, one-time tests. The changes met several objections to the new assessment methods. These alternative assessments:

- ◆ are cumbersome and time-consuming to implement, and therefore, expensive;
- ◆ are not particularly effective as measures of students performance (without attention to the scoring problems; reliability issues);
- ◆ rely frequently on other teachers marking the performances, which does not instill credibility in the approach;
- ◆ sometimes include difficult decisions about criteria (tend to underestimate the assessment of knowledge);
- ◆ emphasize problem-solving while reducing emphasis on learning a body of knowledge, which many people feel is the primary purpose of formal schooling;
- ◆ sample a limited amount of material (limited coverage and some pupils may have read the material or used it before) and pupils sometimes do not actually answer the same questions or carry out the same projects (could result in bias);
- ◆ do not guarantee greater effectiveness in the classroom;¹⁴
- ◆ use extended-response questions often, which appear to put those with poor English skills at a disadvantage and prove problematic for some groups of pupils;
- ◆ measure complex aspects of individuals that sometimes included attitudes and beliefs and other background information about the person producing the product, all of which constitutes an invasion of privacy; and
- ◆ are based on an as-yet infant knowledge base of assessment methodology that sounds less “scientific” than traditional test measurement.

For many development contexts, most of these issues would receive little attention. The major problem is the credibility of the entire new assessment approach. It has not matured enough to mark a credible path for general concerns about educational effectiveness or more specific directions for policy formulation and evaluation. Interestingly, this has emerged as a key issue in the United States as well. As early as 1988, the Congress authorized the development of a national test, a proposal that has not yet been implemented because of possible infringements on the tradition of state and local control over education and related criticisms concerned with the narrowing and standardization of curriculum. Other avenues have been activated and expanded. In 1964, the Exploratory Committee on the Assessment of Progress in Education

¹⁴Linn, R.L., Baker, E.L., and Dunbar, S.B. (1991). *Complex performance-based assessment and validation criteria*. Los Angeles, CA: UCLA Center for Research on Evaluation, Standards, and Student Testing.

(ECAPE)¹⁵ was sponsored by the Carnegie Foundation to design a system to measure “the condition and progress of American education.” ECAPE became operational in 1970 and known as the National Assessment of Education Progress (NAEP), with a shift of emphasis to actual assessments, analysis, and reporting. Beginning in 1990, NAEP included state assessments, and by 1994, Congress removed restrictions to assessment for any unit within states. NAEP is currently a low-stakes assessment program. Congress expanded the role of NAEP in 1984 to avoid the clash with state and local education authorities. But the enthusiasm for some form of national, high-stakes testing has not subsided completely.

President Clinton’s administration in the United States has focused on the setting of national standards, without prescribing the form of assessment. The central problem is that pupils may perform poorly because they have not had the opportunity to learn under local schooling conditions. There is a hesitancy to move into accountability in too strong a manner. Private funding efforts have been more aggressive, and the College Board’s Pacesetter Project and the New Standards Project have strong support among states that include approximately half the pupils in the country.¹⁶ Underlining the strength of standardized testing for credibility, the state of Vermont, which pioneered the use of portfolios, has adopted the combination of its voluntary portfolio program with mandatory state standardized tests to see how well the portfolios are doing in meeting the state’s new academic aspirations.¹⁷ Vermont was the first state to use portfolio assessments as the key indicator of pupil performance (since joined by Kentucky). This move indicates the credibility problem that alternative assessment strategies still have. It’s doubtful that development-oriented countries would move to these alternative methods without checks on their effectiveness, as determined by traditional means (which poses interesting conflicts and technical problems). Even alternative assessments are not the ultimate OLBs!

Complex systems continue to evolve and self-organize. Programs that are appropriate at one time for one country may not be at another time or in another country. Generalizations are possible only on broad dimensions and these can decay or change. The dynamics of education are not comprehensively reflected in testing programs, and the patterns of influence change over time as teachers and pupils adjust to the requirements of new programs and interventions. The claim that good tests will provide useful information on the quality of the education system is too simplistic.¹⁸ And the hope that changing tests in perspective alone will improve the quality of the education system is too optimistic. Therefore, even in prior consideration of the notion of EDI,

¹⁵For details, see Jones, L.V. (1996). A history of national assessment of educational progress and some questions about its future. *Educational Researcher*, 25 (7), 15-22.

¹⁶See Rothman (1995) op. cit., for a more complete story.

¹⁷Manzo, K.K. (1996). Vt. to combine standardized tests with portfolios. *Education Week*, December 4, 1996, 3.

¹⁸See Berliner, D.C., and Biddle, B.J. (1995). *The manufactured crisis*. Reading, MA: Addison-Wesley Publishing Company. Unsophisticated use of test information can distort the implications of test scores for assessments of system quality.

examination-driven instruction, we recognize the limitations of a simple conception of what will be required to design and implement an effective reform strategy using exams.

Primary Leaving Examinations in Uganda

The major purpose of this study is to review practices in testing in Uganda that typify measurement-driven instruction at the national level. Uganda has developed its own strategies for influencing classrooms in primary school from information in its Primary Leaving Examinations (PLE). In particular, the Uganda National Examinations Board (UNEB) has issued three newsletters to schools, distributed information about school performances and suggested teaching strategies for remediation of problem areas identified in the national exams, and has increased the higher-order level skills invoked in its short-answer format¹⁹ exams in the four examined subjects (English, math, science and health, and social studies). The item content has been changing since 1991, with the intention of increasing the cognitive load and consequently, affecting instructional practices.

It is difficult to map precisely the way in which a national examination program may influence classroom dynamics, and there is no clear-cut strategy in Uganda except for the faith in the power of exams. For that reason, our approach is not tightly rationalized for specific connections between information available to schools and impacts. The focus is on understandings and intentions. That is, if schools understand the general nature of the examinations and the intents of the UNEB, then teachers and pupils will take steps to alter instructional practice or at least understand that they could take such steps. Schools may understand the UNEB intents, organize their programs around their intentions to meet the new requirements in the exams, and fail. The failure may be precisely due to a literal interpretation of the requirements rather than dynamic changes in the flow of instructional activities. Teachers may be unable or unwilling to introduce the dynamic changes necessary to alter the traditional attractors for information-dissemination-style classrooms. There is a large step from intention to actuality in affecting classroom dynamics.

The depth of change is illustrated in Cohen's study²⁰ of a second-grade teacher in California who enthusiastically endorsed reform ideas of the state to teach mathematics in terms of problem solving and communication rather than calculation. This teacher did everything possible, attending workshops, reading literature, and practicing new methods. When Cohen observed her classes, he found that her actual instructional methods and practices were much as they used to be! He reasoned that classroom participants "must un-learn much of what they

¹⁹Uganda has not changed the format of PLE examinations in this reform; they've changed perspective within the position of one format. All examinations are short-answer. They believe that this format accommodates the content and skill levels they wish to tap in the examinations. They also believe that the power of the exam will permit a change in perspective even without other changes. Actually in this belief, their resources permit no other option. They hedge their bets with some training courses and newsletters.

²⁰Cohen, D.K. (1991). Revolution in one classroom (or then again, was it?), *American Educator*, 15 (2), 16-23 and 44-48.

know, whether they are second graders or veteran teachers.” Even apparent adoption and implementation is not enough to alter the perspective of a classroom and change its essential dynamics for a different kind of knowing and acting.

Modern reform is dramatic. Do we anticipate that high-stakes examinations can turn the tide alone? or be a piece of some strategy? The evidence from the American scene is clear-cut: exams can change things, though not necessarily what is targeted, and the effects change over time. Political considerations overwhelm the process. Uganda’s reform did not take into consideration these factors. It is a simple reform, and the practice of educational debate is not widely exercised. The question is raised in a context where teachers are unlikely to grasp quickly the depth of change intended, and the politics are focused on regional differences and not educational practices. Uganda is unlikely to have the debates of Littleton, Colorado, or the dramatic perspective alterations like those in Vermont and Kentucky. Uganda changed the perspective, and tried to reach the teachers in formal ways to alert them to the changes. The changes have been gradual. The context presents a picture of what happens with an OLB reform, and helps us characterize the reasons for examination influences and the ways the influences play out in people’s lives. The picture is humbling and sobering.

Teaching and Learning

Teaching is characterized as an *intentional activity* that seeks to promote learning in the formal classroom setting (Pearson, 1989²¹). In evaluative terms, good teaching is simply the achievement of learning by others (although the teacher may learn in the process as well). Teaching does not “produce” learning; it affords opportunities to learn. The nature of learning and the means employed to encourage it are to considerable degree dependent on social and cultural influences. So, what counts as learning will vary by context, but the essence of teaching is the intention to promote learning. How teaching is carried out counts as the practice of teaching, but the action language employed to talk about this process, which sufficiently describes individual teacher actions, is not sufficient to explain the practice of teaching. Teaching is a family of related activities that are tied together by intention rather than just collections of behaviors. The intention explicates what is afforded by the experience. Moral considerations distinguish teaching from indoctrination; empirical findings uncover actions that are more effective than others in particular situations; and to understand teaching, it is necessary to understand the beliefs and intentions of the teacher (and over the long haul, perhaps the pupil as well since the pupil’s intentions can subvert or support those of the teacher).

The following four attributes are associated with good intentions to promote learning (see David Perkins, 1992²²): (1) clarity of intentions, (2) opportunity for thoughtful practice, (3) provision of informative feedback, and (4) use of reflection, and strong intrinsic or extrinsic motivation. These attributes are the targets for many reforms, in most cases indirectly through the

²¹Pearson, A. (1989). *The teacher: Theory and practice in teacher education*. NY: Routledge.

²²Perkins, D. (1992). *Smart schools: From training memories to educating minds*. NY: The Free Press.

rationalization of the system. These categories are not mutually exclusive but present the array of actions that frame the effective classroom.

Pupils must “see” the teacher’s intentions and beliefs in order to understand what is going on in school. The teacher’s intentions define the classroom situation facing pupils. The pupils must consider the actions of the teacher as purposive so that they can make sense of how they can view themselves as pupils. Even the young child in an early grade must understand the context in order to actively participate. The pupils must *trust* the teacher. In a sense, they *submit* to the intentions of the teacher as well as to what is being taught (subject area and material delimit further the situation and convey their own intentions), even if considerable freedom is available for pupil choice and participation. Learning, of course, does not require teaching. Teaching is an *INUS condition* (that is, an insufficient but non-redundant part of an unnecessary but sufficient condition) of learning so that it can be construed as causal in this sense. There are no strict rules for effective teaching. It is possible, however, to develop prescriptions and have them work, but this does not optimize conditions for each situation (although it may be necessary as a starting point when teachers lack the range and depth of teaching repertoires²³). As the philosopher of education Allen Pearson²⁴ points out: Teaching is largely strategic in its best form, such that effective practice depends upon following strategic rules that contribute towards success in the classroom rather than absolute rules. Teaching is a social activity, carried out amidst a group, sometimes a very large one. The meaning of the classroom is created by the intentions that the teacher and the pupils bring to their situation. These intentions are drawn from the knowledge that each has in coming into the classroom (Williamson’s notion of the contract and subsequent transactions).

The teacher, based on some lesson plan that is related to the prescribed syllabi for the subject areas to be taught, intends that the pupil will learn some particular item or skill or understanding. When the pupil “buys into” this intention and engages the activity, the situation is intentional for both. The lesson, therefore creates an “intentional situation.” The questions that emerge are how smooth is the mutual engagement in the situation and what kind of learning is intended. Here we are primarily concerned with the dynamics of the situation, but consider those dynamics in the context of content. Different subject areas have evolved different “cultures,” and their dynamics may also differ.

²³Careful presentation of specific recommendations for teacher actions help to clarify the teaching role to those who are just learning, themselves, about the possibilities for classroom actions. This does not preclude later loosening of the prescriptions (or different kinds of clarifications). The intent is to provide some structure to the instructional program. In some cases, for example Instructional Systems Design, the prescriptions are more forceful. The intents are programmed by design. This is the rationale of the USAID-funded *Basic Education Support Project* in Namibia.

²⁴Pearson, A. (1989). *The teacher: Theory and practice in teacher education*. NY: Routledge.

There were 23 items used in the assessments of classrooms in Uganda.²⁵ See Table 3 for the specific items. This may appear to be a small number of scales to characterize an entire lesson. Each item is complex, laden with content and encompassing a number of actions, but remains independent of resources and facilities. Learning, in other words, can take place under many particular circumstances and kinds of behaviors. These items assess the pedagogical process within the context of a subject area, and this process is the main thing that should be affected by the testing intervention if a change in perspective will be possible over the long-term. These items characterize the minimal qualities that must be present for a *smart classroom* (see Perkins, 1992), one that would represent the change in perspective to a thinking agenda.

Case Studies of Schools

Sixteen schools in two districts constituted the sample. The school selections were based on a previous sampling due to the Improving Educational Quality (IEQ) Project,²⁶ where schools were earlier studied at lower grades than Primary 7 (the year of preparation for the Primary Leaving Examinations). There were still many problems with armed incursions and conflict, so the study was restricted to the two central districts. Uganda has a long way to go to recover from the many years of neglect and devastation. The existence and organization of schooling is testimony to the importance attributed to schools. In institutional terms, this is an extraordinarily successful system, a system which attracts resources and attention even during difficult historical times for the society. The power of examinations is quite evident, and stories abound of conflict curtailment for the administration of the all-important examinations. These schools reflect the intact schooling system of Uganda. They are particularly relevant for this study of the influences of examinations in developing contexts. The power of examinations has been studied in resource-rich contexts, but has received only scant attention in contexts where few resources are available or likely to be available. There is no intent to generalize to all of Ugandan schools. We treat the

²⁵Each item has a rubric (set of criteria) associated with it to anchor the levels for rating. The ratings were all carried out by an American female, who had two Masters degrees (education from Harvard and health from Johns Hopkins). She had extensive overseas experience. Even with these impressive credentials she experienced difficulty due to the complexity of the questionnaire and the simplicity and lack of variation of the classrooms. She improvised by using higher scores for what variation she did find. This adjustment of the rubrics may have changed the nature of the scales over the course of the observations (she reported stability after the first three cases). This bias proved unimportant because mean levels were still low. The classrooms and classes have differences, but few of these really relate to the process of education. As indicated by Fuller and Snyder, 1991 (*Comparative Education Review*, 35, 274-294), there is a great deal of homogeneity in southern African classrooms, but as illustrated in Snyder, 1990 (Snyder and Ramatsui, *Curriculum in the classroom*. Macmillan), the variation in the dynamics of teacher worklife is telling and predictive of examination performance. There is more going on in the classroom than we yet understand.

²⁶The initial plan was to include data from the IEQ in this study. Access to the final data never materialized. This was unfortunate. Whereas this study focused on Primary 7, the seventh year of primary school, the IEQ study had data on these same schools at other levels of the program. These data included classroom observations that would tell the stories of preparation, prior to the last year, for the high-stakes examinations. When available, the IEQ reports from Uganda should be consulted for a more complete picture of schooling practices in these schools.

schools as case studies, even when using collective information to increase the range of experiences. The study says less about Ugandan education than about the influence of examinations in particular schools, which operate under very difficult circumstances and for whom the Primary Leaving Examinations are particularly high-stakes for their resident pupils.

Within these 16 schools, 60 teachers²⁷ and 15 headteachers were interviewed.²⁸ Thirty-one classrooms at the Primary 7 level were observed. In 15 schools, about 34 groups of pupils were interviewed, with about 4 pupils in each group on average. In 8 schools, 10 interview sessions were held with community members and parents, with about 4 persons in each session on average. One of these interview sessions is on tape. For 12 of the 16 schools, the PLE from the previous year were available. In this PLE, pupils had been asked, "Should the PLE be abolished?" They wrote essays on the topic. The data set therefore consists of:

- ◆ PLE essays from last year's pupils of their attitudes towards the PLE;
- ◆ Interviews, structured and free-flowing, of teachers,²⁹ headteachers, pupils, community members, and parents; and
- ◆ Classroom observations of current pupils preparing for the PLE.

Many of the pupil, community, and parent interviews remain unanalyzed. The data collected from these few schools is enormous. The open-ended nature of the data provided a rich amount of material. The field study was stopped in September 1996, prior to the PLE examinations of that year. No funds exist to collect those data for the classrooms observed, so references to influences of examinations are limited to prior school performances and verbal reports. It is the prior results and the UNEB changes that should have affected these classrooms so this is not a weakness in that sense. However, we have no way of gauging the relative success of various classroom dynamics on the "new" examinations, nor can we investigate the relationships between the classroom observations and performances for that particular year.

The original proposal for this study entailed a time-series analysis of classroom changes during different segments of the final primary school year, where the PLE might present differential influence. So this study is limited to mid-year activities, which are building to examination preparation. The impression is that the PLE exerts even greater pressure and influence over the next months of classroom work. The drama of examination influence was not complete during our interviews but the emotion was not lacking. Individuals, whether parents, teachers, or pupils, took the PLE very seriously.

²⁷One of these teachers was also a headteacher/principal but chose to categorize himself as a teacher.

²⁸The interview protocols were trialed in Lesotho (see the report by Khati and Bancroft, 1996). This preliminary study of the influence of examinations in Lesotho was funded by the Center for Higher Education and International Programs of Ohio University. Ms. Khati and Ms. Bancroft received course credit for their report.

²⁹See Volume 2 for the categorization of many quotes from pupils and teachers.

Influence on Pupils?

The idea of examinations is to focus performance and expectations, motivate individuals toward explicit objectives, and provide feedback on the quality of performance. How scores on examinations are used determines their ultimate purpose, and that use sometimes differs from the development assumptions of the examinations. For example, the PLE is a set of “leaving examinations,” which would imply that they assess the accomplishments of pupils as they leave the primary education level. Some countries include the notion of a certificate linked to passing the examinations to underline their role in accreditation of the graduation. However, in Uganda, like most other places, these examinations are actually aimed at the selection of schooling participants to the next level, under conditions of restricted positions at this higher level. Schooling in Uganda is provided for all pupils through seven years,³⁰ and then is limited to those who “pass” the examinations at the next levels. Examinations serve as public accreditation, with or without technical accompaniment, but their real function is selection. This has posed a dilemma for test developers for some time. Achievement tests perform the assessment of the past; aptitude tests predict future performance. These technical differences, however, prove to be illusive in practice. It would be difficult to categorize a single item in either form without additional information other than the item. Fortunately, one of the better predictors of future performance is past performance, so the distinction has little utility and few give thought to it in practice.³¹ The importance lies in how we conceptualize the examinations and what influence we encourage for instruction, either explicitly or implicitly.

Pupils are aware of the functions of examinations in terms of certification and selection, even if implicitly so. Table 4 shows the “vote” of those taking the examinations on whether or not to abolish the PLE. For those who attribute some effectiveness to the PLE, they mention in their essays the role of the PLE in marking completion of the level, motivation of pupils and teachers, certification of performance, selection and prediction of future success, and the evaluation of teaching. Many pupils, in fact, see the examinations as a key component to their notion of standards. These independent examinations promote competition, provide recognition

³⁰This is not precise. Only those children whose parents pay school fees go to school, even primary school. Lots of children do not go to school at all, and many drop out for lack of money.

³¹This isn't entirely true. Schwarz and Krug, many years ago during their collaboration at the American Institutes for Research (AIR), suggested the combination of achievement tests and aptitude tests for use in Nigeria. The Internationally Developed (ID) tests emerged, based on the Flanagan Aptitude Classification Tests (FACT). ID tests were used in many countries in Africa. The idea was to use the achievement portion of the examination set to provide a certificate for graduation from the schooling level; whereas, the aptitude portion could be used to be sure not to miss a future achiever at the next level.

Also, we have already mentioned the debate about the purpose of exams to measure exactly what was taught versus future applications, which are not content dependent (a version of aptitudes). In Africa, this was illustrated in the development of the Swaziland Educational Development Scales: Cognitive Tests for Standard II (Primary) by the first author and many others for use at the fourth year of schooling. The series was never used, however, except in successful trials.

for outstanding schools, prevent extensive corruption and local inequality, and establish the standards for national education. As an example (see Volume 2 for details on all quotations),

[The PLE] shows the standard of the school, the teaching of the teacher, whether they are lazy or hard working. Not only those, it shows how the children are they, are they good? Or are they weak?

...how can someone tell you have finished the primary when you're going in secondary schools. Or how can your teachers know that you have understood what you have been taught in primary.

If this examination was not there, it would mean that all pupils who are in primary seven would just go to secondary. And if this is done, secondary schools would receive pupils who are not fit for secondary.

The vote on abolishment of the PLE is close, and the number of unclear responses could turn the tide in either direction. These results are confused by the fact that few pupils, as we learned from the field trips to schools, actually understand what “abolish” means. We used the same term in our questions and were met with puzzled looks. In reading the texts, it is clear that in many cases someone has “clarified” the meaning, but not always at the level of full understanding.³² There is commonality in the initial frame, and then pupils take off on their own paths (well, many don’t answer at all anyway or remain unclear). At times, the vote and the text were not consistent. Perhaps that’s not unusual!

Pupils more generally talked about the dysfunctions of examinations in their scripts. A common phrase in these texts and in our discussions with pupils was *Exam Fever*, which has become our metaphor too for the pupil responses. We categorized the scripts into references to physical, mental/emotional, social/family, teaching/learning, and education system dysfunctions. These are not all negative. For example, one pupil points out that failure in these examinations may lead to death to a child (for failure to read expiration dates on medicine). All kinds of events, attributes, and futures are ascribed to these examinations. They clearly loom large in the lives and futures of these children. The individual details contain many novel and different ideas about the influence of examinations.

...some candidates can get mad due to reasoning so much, and others, their friends decide to give them poison in order to win them. In addition to that, some candidates die because of their friends going to witch doctors to give them herbs in order to kill their friends.

...they caused sickness to pupils. If they know about them they started panicking and panicking. Some times [the PLE] can cause death.

PLE should be abolished because the child can fail the exams and their parent will beat him or her.

Primary leaving examinations makes a pupil to panic. From there a pupil develops what we call examination fever and diarrhea.

³²Our “voting” tallies should be correct because we used the full text, not just the child’s use of the word abolish.

The academic roles and ultimate functions of examinations are understood by some pupils in Uganda, and these texts appear to be of reasonable quality and from very few schools. For most pupils, the examinations have a relationship to some dysfunction.³³ The high-level meanings of examinations are lost or unimportant to those most affected by the events. They see many other consequences and associations that play a concrete role in their lives. We could take this concreteness as a failure of the education program to promote skills to take on and present the conceptual issues of examinations or we could attribute these personal responses to the particular importance (high-stakes) of examinations to each child (or both of course). There is little doubt that these examinations are viewed as high-stakes.

On my side, I say that this examination should not be abolished because it shows the standard of the school. Since this examination is set on international level, from their results, the schools' standard is seen which promotes hardworking.

Still primary leaving exams should be done because it shows the standard of the school nation wide. There are other exams which are set by schools which are not of national level.

A person can read books from morning up to night and from night up to morning, but reaching the exams you fail to see any thing they have brought which you read. You can develop a fever at that very time and you waste all seven years.

When we are in primary seven, we overstudy. We study day and night. In some schools, they study even on weekends instead of going to worship God. If we don't worship God who helps us in our exams, why should we continue with such examinations?

Some pupils fear them and they got examination fever in their stomachs but they are good if you do not get examination fever.

Primary leaving examination should be abolished because some children get brain damage. Children fear to fail because their parents will complain so they keep on reading their books without giving time to their brain to relax and they end up running mad.

Primary teachers recognize and can largely postdict the important issues mentioned by the pupils. They understand that pupils fear the examinations, suffer physical and emotional stress, worry about many aspects of preparation and fairness, need competition and external motivation, and recognize many consequences to the examination process and results. The pupil essays appear to be valid reflections of their views, as possibly expressed in the context of their schooling or as recognized by those sharing the schooling experience. The impression from reading many texts and talking to pupils in groups at schools is that these examinations arouse diverse and deep feelings and have many impacts on personal lives. We tend to focus on the technical aspects of examination programs and forget the human consequences of their influence.

PLE makes the students very tired, so tired that they will not think straight while they take the exam. Most of them do not do well in PLE. If those who pass do not get places

³³We expected, from an American perspective, that the examination context would encourage positive responses about examinations. That doesn't appear to be the bias in Uganda!

for secondary, they end up feeling that it was PLE which stopped or blocked their chances.

They think it is overburdening their heads. Some spend sleepless nights. Others said that the exams failed them.

Because students fear examination. If PLE is abolished they would automatically go to senior secondary school.

Students feel this way because they feel it is a hard hurdle to go over, in that it gives them a hard time in life for a short time.

These are high-stakes examinations with many side effects due to the importance of the contingencies. The question is: Can the examinations can serve their educational function well while they also exhibit so many adverse effects? Can we control the value of the experience? More importantly, can we ensure that no damage is done? From the pupils' perspective, it appears that there are diverse views of what's happening in these high-stakes examinations, why, and the consequences. And it doesn't seem we can guarantee no damage to individuals. Not everyone is served well by this experience.

PLE creates fear among the learners at this stage. The way of PLE needs to be reviewed in order to allow the children to take PLE in a more relaxed way. Children take a lot of their time preparing for PLE. They have no free weekends at all because of preparing for PLE. The children end up disliking PLE because of the pressure it bears upon them. Students' attitude toward the PLE has convinced me that PLE should be abolished.

Examinations just test very little of the cognitive part of the child, in that some children who do well in the classroom do get affected during exams and perform poorly, much beyond one's expectation.

Very unfair way to judge students' performance over seven years' work and done in two days. If a child falls sick and ends up not doing well, there is no alternative for her or him. Seven years' work is compressed into 100 question items. Obviously there is the possibility to not do well, especially in those areas the child will not have stressed. For example, the students who fall sick during PLE end up doing poorly. In many cases the candidates get frustrated when the language is above their heads.

It has brought a lot of malpractices in education. We should have continuous assessment.

Influence on Teachers?

In the consideration of their own positions on the disadvantages of the PLE, teacher responses talk about their concerns about exam orientation, role of fear and stress, question of fairness, difficulty of language, occurrence of cheating, malpractices, and stealing of exams, impact of failure, expense of the process, disadvantages to poor and rural children, general difficulty of the exams, and drop-outs due to exams. High-stakes examinations pose many perplexities to educators everywhere.

The one-time national exam leaves out many students with a potential to do well. Thus the exam wastes away good students. The exam results have led to a very high rate of dropouts because each time they sit for the PLE, large numbers fail to qualify for

secondary level. The national exam has led to a bookish system of education where students and teachers worry more about learning to do well in the examination over everything else.

It can lead a brighter candidate to fail. It may be the first time to sit for the papers. They fail because they believe the examination is difficult. They can get exam fever.

Not all parents or students are economically the same. All schools do not benefit the same way. Some lack materials and equipment. Some have poor buildings. Trained teachers are not evenly distributed. For example, rural schools tend to get poorer trained teachers than in urban areas.

In one way or the other, it discourages some children who feel that the PLE is very hard, whether they study hard or not. Some may end up going to poorer schools. I find that it accelerates the rate of dropouts, especially the post P7 level. Compared to the current situation in the system of education, planners and administrators tend to stress passing P7 instead of looking at a rounded education with culture, etc. The PLE has resulted in malpractices—drilling, rigging exams, and corrupting the UNEB officials, for example the invigilators.

The teachers have a great deal of compassion for their students and seemingly appreciate their anxieties, but they are supporters of the examinations and widely recognize the advantages of a national examination. Their attitudes and responses indicate a certain professional fervor for the retention of exams, and thus we selected this as our second metaphor. They see the exam as a fair way of judging pupil differences, guaranteeing a national system, providing evaluation of instructional process, motivating teachers and pupils, instilling the notion of competition, and providing a uniform system for individual planning and selection.

The PLE national examination acts as a yard stick used to promote children from primary to secondary level. The PLE also judges how well teachers perform: usually the pupil's poor performance is attributed to teachers' poor performance—however, when pupils do well, the teachers are praised and taken as very good teachers.

We can tell the levels of understanding among our students. It is also a measure to enable the teacher to compare with other schools in other places nationally.

Makes pupils in the whole country do the same exams so no cheating [occurs] when they are being promoted. [PLE] makes non-performers repeat the class.

There are many children for few places in secondary school. So PLE is good in determining the few to enter the next stage (secondary). Government cannot take all PLE students. PLE encourages country-wide good performance.

To guarantee the national system, i.e., when we do the national examinations, this testifies how well the teachers have gone through that particular syllabus. They give teachers a challenge as they look at the performance of their children and other pupils, nationally. Reward a teacher if PLE results are good; make them work harder if they are poor.

The PLE acts as a motivator of students and teachers. It helps assess progress of students and their school in general. School culture—the national exam helps us to tell if

curriculum suits our pupils and their needs. PLE also helps us screen children according to their abilities for vocational schools and academic schools.

Exam influence depends upon the recognition of change-in-perspective by the teachers. UNEB has tried to convey the move of the PLE from solely being a test of knowledge to one of higher level cognition. Changes take time. UNEB began its changes in 1991, and the wide diffusion of information about the changes has been underway by the time this study began. The UNEB alterations have been done slowly and efficiently. This was not a high-profile modification of perspective, except for its link to examinations. UNEB counted on the attention teachers give to past exams, information presented in newsletters and workshops, and general diffusion of the changes.

[Since] 1994 and 1995 there are [fewer] grammar [questions] now. Now there are more [questions emphasizing] understanding. After the final examination, there's a relation between what they've learned in class. More thinking and reasoning is asked by the questions. Questions involve answering in full sentences. Not as many, if any, questions involve just one word for the answers. Tricks have made us more exam-oriented.

The particular changes I have noticed are that the type of questions being asked are different from those in past years.

Things are now becoming new. They want to test the capability of young Ugandans today.

Questioning techniques are changing. They are using different words. Some years back students were not asked the reason questions or the why questions. Now a variety of techniques is used.

The setting of examinations now is mainly based on questions based on thinking. No more setting [of exam questions] directly from books. These people have always tended to go beyond the syllabus when setting the questions.

Last year they asked questions that require reasoning. Before, questions were about what you remember and didn't need reasoning.

The mode of questioning gives pupils to reason....There are no direct questions, but only reasoning.

Before there were objective questions, but today it requires thinking. As one teaches, the pupil himself has to write the materials that is in his head. He has to write his own information which he knows, rather than arranging answers.

The way of setting examinations more especially. That is, before 1994, exams were usually on facts. Since 1994, setting has focused [on] and tested much of the reasoning of the child, more especially in the areas of science and social studies.

Many teachers mentioned the relevant changes and exhibited a good understanding of what was going on. From the responses, we learned that teachers took awhile to notice and be aware of the nature of changes. Even changing exams is not a "quick fix." Still, UNEB, under difficult circumstances and limited resources, has been a success in developing this change-in-perspective. They merit considerable admiration.

Some Quantitative Details

The classroom observation questions used in the Ugandan study were categorized as clarity, thoughtful practice, motivation, and feedback (see Table 3). The protocol was first developed and trialed by district resource teachers in Lesotho in 12 rural schools. Teachers were each given classroom profiles to discuss their classroom approaches with each other. The picture below shows the teachers at one of the remote primary schools reviewing the profile on key items from the classroom observations. Because the observations were carried out by the district resource teachers, the results were considered helpful rather than evaluative. Based on the data from Lesotho and discussions with teachers and district resource teachers, some modifications were made to the Lesotho version of the classroom observation items for use in Uganda.³⁴



Figure 3. Teachers review classroom observation information.

Thirty-one classroom observations represent a significant amount of field work. For multivariate analyses, however, this is a small number of units. Nevertheless, we checked the internal consistency and the structure of the scales. Alpha reliability coefficients³⁵ for the four scales showed consistency in the scales: 0.71 for clarity items ($i=4$), 0.79 for thoughtful practice items ($i=4$), 0.77 for motivation items ($i=5$), and feedback only had two items but had an alpha

³⁴The instrument was developed as part of the *Lesotho Primary Education Project (PEP)*, under the auspices of Ohio University. Although the principal architect of the approach is Wes Snyder, the actual wording of questions and rubrics were worked out with inspectorate personnel in workshops with Lynn Evans. The current version is another modification based on the Lesotho experience. See Snyder, C.W., Jr., Evans, L, Leherr, K., and Makoele, M. (1996). *School-based study. Progress report 2*. Maseru, Lesotho: PEP.

³⁵Interrater reliabilities were attempted in the Lesotho study but are not reported because we believe that the raters collaborated in many instances.

coefficient of 0.80. These four scales were analyzed by Principal Components Analysis. Feedback and Clarity correlated 0.51, and Thoughtful Practice and Motivation correlated 0.67. Accordingly, two components accounted for 80% of the total variance. These relationships make good pedagogical sense. Feedback is an important complement to clear communication and probably aids in understanding the intents of the lesson. Thoughtful practice is a good complement to attempts to motivate pupils. If the practice is thoughtful, then pupils begin to understand the importance of the activities, which in turn helps to motivate them as relevant for them.

This was not a quantitative study, in the sense of large samples for the estimation of population parameters for Ugandan education. For each of the scales a principal components analysis was calculated to examine the structure of item relationships in terms of total variance (see Table 5). For clarity, Item C6 was not well accounted for by either of the two components, and C2 was relatively unrelated to the other four items. For thoughtful practice, Items TP3 and 4 are better accounted for by the second component, while the other items contribute to the variance of the first component. For motivation, Items M1-5 mark the first component, Items M6, 7, and 9, the second component, and Item M8 marks the third component. In the clarity solution, the amount of total variance accounted for is 54 percent, and about 70 percent for the other two solutions for thoughtful practice and motivation. The lower variance in the clarity solution is due to the lack of fit for Item C6. Alpha coefficients were calculated, as above, on the items that swarmed around the first component. The categories of items were not developed to measure unidimensional features of instruction so this analysis is more descriptive of the item relationships.

To the extent that these items get at the notions of Perkins' minimal requirements for a thinking classroom, they provide an important picture of classroom dynamics in Uganda. The means and standard deviations are provided in Table 5 for all items. Scales ranged from a low value of one to a high value of eight and were behaviorally anchored at four equally-spaced locations. We can characterize these classes as having objectives present, with moderate use of elaboration and examples, moderately accurate and thorough knowledge of content, and moderate coherence to the lesson contents. There were few supportive activities on average and no flow to minimal flow to the process. In terms of thoughtful practice, there were minimal opportunities, and learning focused on the understanding of facts, on average. Some preparation of the lessons was evident, some pupils participated actively in the lesson, and the teacher did ask questions of some individual pupils. The lessons entailed, on average, a single teaching strategy that enabled some pupils to engage in the learning process, and attracted a moderate percentage of the pupils to stay on task. Teachers communicate moderate interest in the subjects they teach, with enjoyment evident in some pupils. There is almost no pupil initiative or creativity displayed. The complexity of the different points and perspectives is minimal in the lesson, and yet the difficulty of the lessons challenges a moderate number of pupils. Feedback is infrequent and simplistic. This presents, on average, a picture of very traditional classrooms with teacher-centered instruction and limited, if any, thinking or opportunity to develop thinking habits or skills. The average classroom is not what the examination reform team had in mind. Clearly, these pupils are ill-prepared for the "new" examination. Since UNEB has been sensitive to the

context, the PLE is not a totally revamped set of assessments, so items remain to sort through the factual knowledge holders. Average values on these classroom observation items are low in general and not approaching the levels expected of a reformed system. Although teachers report awareness and influence, it is not even somewhat evident in the classroom. There are very few (maybe one or two) *smart* classrooms.³⁶

Influence on Classrooms from the Teachers' Perspectives?

Classroom teachers were asked how much their efforts in class were aimed at the Primary Leaving Examinations (PLE), how well their pupils learn valuable information in general (Information), or just trying to do their best teaching (Trying). Thurstone scaling of their rankings indicated that Information was preferred over PLE approximately the same as PLE was preferred over Trying. From side comments, "just trying" was seen as pejorative by some teachers. Ironically, those teachers selecting Trying as the most important had the highest ratings from the classroom observations in three of the four scales (Clarity, Motivation, and Feedback). Teaching-centered classrooms probably have more structure, which provides some basis for good instruction and learning.

In response to "how has the PLE affected you?," teachers attributed many things to the exams. Mostly, the exams make them work harder, spend more time with pupils, even after hours, and require them to cover more material in the syllabi. The PLE focuses their work and that of the pupils. The PLE becomes the defining event for the activities leading up to the examination administration.

I am forced to work tirelessly because I wish to keep the spirit of my students being happy. So each year I work twice as hard as the previous year.

[The PLE] overburdens me. The rate of teaching increases, where parents want me to begin teaching at 7:00 and end at 6:00, so that we can compete with other schools. They don't even consider the problems we have in this school. They think that teaching a lot will make the children pass, and don't consider the environment or materials. I have a headache every day.

The PLE has been a check point for me. When students fail, it means you have not covered the syllabus. In P7, we have more work than in other classes. Like now, we are giving premock exams again and again. In P5 and P6, we do tests at the end of the month only. In P7, it's almost daily and weekly that we give tests. In P7, we lead them in group discussions and then let them work on their own so that they can explore for themselves. In P5 and P6, we keep them together so that we can guide them.

³⁶This is particularly troubling because teachers did their "best" for our observer. Sampling over more occasions may reveal an even bleaker picture for the impact of the reform. The *Quality Link* (Winter, 1997, 10) of the *Improving Educational Quality* (IEQ) Project indicated: "First-cycle research findings confirm that schools are deficient in all areas that contribute to school effectiveness including adequate curriculum standards, instructional supervision, and material and financial resources." But, "researchers discovered that students performed markedly better at level P6 because the best teachers were often assigned to the P6 level to prepare pupils for the PLE exam." We probably saw the best teachers.

Teachers are encouraged to work hard to get good results and performance. Teachers read more extensively to be knowledgeable. When [PLE] results are good, your relations with parents is very good, but if results are poor, you have parents hating you. [The PLE] give me morale to work harder to have more [pupils] pass. [You get an] uneasy feeling if students don't perform and you may lose your job if they perform badly.

[The PLE] makes teachers improve on their teaching methods. It makes them follow the syllabus more closely. [The PLE] can result in feeling that if I do well, I will be promoted and that if I do badly, I may be demoted or transferred. Poor performance of students affects me being removed from or transferred from school. I may lose [my job] or decide to quit teaching altogether. I feel like improving both my teaching and administration. [The PLE] encourages me to buy my teachers some teaching aids. Teachers are not quite adequate in teaching the students, since some of the P5 and P6 [teachers] are not trained. [Student motivation] has not changed. Teachers are in class very early in the morning to help students. Students are eager to come to school early.

For the initial question about the PLE impact, only one teacher mentioned a new teaching methodology.

Teachers also work hard to cover work which they believe will help their students to pass well. Teachers teach in order to be recognized for performing well in their school. I have to ensure that I have in place an effective monitoring arrangement. I have to assist my teachers and students in order for them to produce good work. When students perform well the whole community has a lot of respect for him or her [the headteacher], but when they do poorly, the opposite is the case. Teachers have resorted to the discovery type of teaching and they read widely and encourage their students to explore further areas to come up with new discoveries.

In follow-up questions seeking more specific effects of the PLE, one of the highest rated teachers in our classroom observations said this:

In a classroom, teachers don't concentrate on facts alone, but on how facts can be used after school. Teachers have changed their methods of teaching. Because when you teach somethings' advantages and disadvantages and when it can be applied, you have to teach that way because of PLE changes. Teachers have turned out to be more exploitive [exploratory], looking for other materials and facts to give pupils. They don't concentrate on the content. They relate what's in the tests to their experiences. They visit other schools that are performing well. [The PLE] leads me to explore more than just giving the facts to the child. [For example, before, a question on the PLE might be] "What is a district of Uganda?" [Now the question would say] "In what ways is Kampala important to Uganda?"

Another teacher points out the discrepancy for children between prior grades and P7:

Right from P1, teachers pump children only [with] facts. When it comes to P7, they're forced to orient pupils to thinking. There is a discrepancy. Facts in P1 through P6 years, [but] in P7, thinking is needed. There is difficulty in orienting P7 pupils to thinking. When children are given work, they do very poorly in terms of thinking. They're not creative. They don't think and create things on their own. They get a lot of problems involving thinking. They become confused. But if we give direct questions needing direct answers, the pupils are able to do them. As a P7 teacher, I have to encourage my

children to think on their own. I have to keep giving them comprehensive questions which make them think.

The PLE reform does not dictate teaching methodology. So teachers can adopt various approaches to preparation for the exams. Another highly rated teacher in our classroom observations points out the differential practices:

We teachers have different ways of teaching. Some teachers do not care about those questioning techniques. So if children do not practice them during the course, they fail exams. Teachers don't change with the change. Students have to practice so that they get familiar with the kinds of questions. They [the changes in the PLE] don't affect me much. For me, I can change with the change. I've now adopted those questioning techniques. I like it.

The direct influence of PLE on subject coverage is questioned by this teacher, but s/he admits to the possible improvements in teaching methods.

Teaching in classrooms has now adopted the PLE model. Other subjects are being neglected. Too much time is spent teaching PLE subjects. The way UNEB asks questions is letting teachers follow techniques. The way it was done in the past is ignored. Now [we are] exam-oriented. Teaching is stressing reviews of all past papers and not for understanding. Some teachers have improved their teaching and have become more resourceful. [There is] stress [from] using discovery methods or [other] new approaches.

And there are different points of view about exam orientation, its advantages and disadvantages.

Now teachers do teach extensively. They no longer drill for the sake of passing PLE exams. Teachers now teach more for understanding and applying what they [children] have learned.

When these changes came, I started noticing that the teachers ignore the syllabus as such and spend a lot of time drilling pupils just to do well on the PLE. I have found that the PLE has encouraged or forced teachers to teach and work much harder than before in every subject area that is involved with PLE. [The PLE] has encouraged the students to think and reason out widely. Teachers now spend more time drilling for doing well on the PLE and not teaching for life.

Most of the teachers and headteachers interviewed reported changes due to the PLE reform (see Table 6 for changes noted by teachers). On a single occasion classroom observation (which was unannounced), we could not detect the differences reported. The classrooms appeared much as we might expect without the reform. Perhaps, the teaching methods will appear more regularly over time; or they may be good intentions expressed in interviews and cast as if they were actions; or they may not be fully routinized in the repertoires of these teachers. Another concern is that teachers-being-observed may believe that *they* are being observed and resort to teacher-centered methods to demonstrate *their* skills in direct instruction.³⁷ However,

³⁷This was the advantage of the time-series design of the original study. Future studies should examine classrooms over time (as per the Botswana Junior Secondary Education Improvement Project studies, particularly the work of Pat Rowell and Bob Prophet in C.W. Snyder, Jr. and P.T. Ramatsui (Eds.). (1990). *Curriculum in the*

rhetoric does change before action so perhaps we are talking to many adopters who are trying some things and honing their skills in new methods, but haven't yet reached a level of confidence with their "performances." Clearly, the reform has changed many teachers. Perhaps action will follow. This is where good support on how to change would be welcomed and effective.

Technical Limitations to the Use of Examinations as a Lever for Reform

The results of this survey of particular schools in Uganda point to the many dimensions of a strategy that first appears as an OLB. We can categorize the larger results using House's clever rendition³⁸ of Williamson's transaction approach to markets and firms,³⁹ called the *Organization Failures Framework* (1975) and later covered in the general consideration of transaction cost economics (1985). In terms of House's application of transaction cost economics to education reform (1996), examinations as a reform device fall short of required technical attributes.

House suggests three dimensions from Williamson's 1985 version of transaction cost economics: *bounded rationality*, *opportunism*, and *asset specificity*. Teachers and pupils enter the classroom operating under certain presumptions, similar to informal contractual understandings. The school takes on a certain form, and this form reflects the expected transactions among people at the micro level, in the classroom, and includes the governmental structures of the school administration. Reforms impinge on these considerations and conditions, possibly altering the "contracts." Testing was not specifically addressed by House. Tests, however, are key components of the contract. Pupils study certain material so that they are prepared for these high-stakes examinations, and teachers cover syllabi of material to make sure that pupils have had an opportunity to learn. The contractual opportunity to learn is frequently broken in development contexts. But we can also invalidate the contract if we assess on material or by processes that children cannot fathom from their preparation, and this can happen in examination reforms. House's framework provides a glimpse into the UNEB examination-driven instruction reform.

As we found in Uganda, teachers are subject to *bounded rationality*⁴⁰—teachers cannot necessarily adjust to examination changes; even if they understand the examination requirements at a cognitive level or an emotional level, they may be able to do very little in making the

classroom: Context of change in Botswana's junior secondary instructional programme. Botswana: Macmillan.

³⁸House, E. (1996). A framework for appraising educational reforms, *Educational Researcher*, 25 (7), 6-14. Unfortunately, we were not aware of this paper before this study began. His framework provides powerful considerations for this analysis.

³⁹Williamson, O.E. (1975). *Markets and hierarchies: Analysis and antitrust implications. A study in the economics of internal organization.* NY: The Free Press. And Williamson, O.E. (1985). *The economic institutions of capitalism: Firms, markets, and relational contracting.* NY: The Free Press.

⁴⁰ "...human behavior is *intendedly* rational, but only *limited* so...bounded rationality—of the behavior of human beings who *satisfice* because they have not the wits to *maximize*." (Simon, H.A., (1976). *Administrative behavior.* 3rd ed. NY: The Free Press, xxviii).

necessary changes in the classroom to affect performance improvements. This was the case with Cohen's second-grade teacher, and this is the case with many teachers in Uganda. Although we cannot directly address the extent of boundedness in this study, most of the teachers and headteachers recognize that change has occurred in the PLE, and only four of these individuals did not recognize the salient change. Unfortunately, no evidence in the classroom observations supports any changes in instructional dynamics as a result of the reform. *Bounded rationality* captures the limited knowledge and rationality of all humans. A successful reform must take this into account. The UNEB reform tried to provide information through various means, but teachers, similar to pupils, are not vessels for information. It cannot be poured in, retained, and applied. In the area of application, the UNEB reform is inadequate. Faith in the power of examinations led to an expectation that change and information would lead to the "right thing." With few resources for support for professional development and other mechanisms to promote the knowledge of the changes and the necessary classroom changes to accommodate the examination requirements, UNEB perhaps had to take the chance that some reform may emerge from individual initiatives. Most of those teachers who recognized the change also claimed that the change affected what happened in the classroom. When one considers the minimal support available for the reform, the cognitive awareness of these teachers is impressive. The only problem is that many children never had a chance to succeed because the old testing contract had been broken while they were prepared for traditional examinations. Perhaps two reform lessons come out of this: exams are powerful influences and many teachers do pay attention to their content and form; but exams are not powerful enough to assure comprehensive reform.

Examinations are naturally subject to *opportunism* or self-interest seeking, particularly in Uganda. Uganda attempts to reduce opportunities for information advantage both *ex ante*, by prior identification of risks, and *ex post*, by standardization of exam procedures. Efforts breakdown because weaknesses are not identified beforehand either by UNEB or revealed by sources in the field or appropriate risk mitigating actions are not taken in the field because of counter incentives. Incentives abound for all concerned to take advantage of weaknesses or oversights in this high-stakes enterprise. Alternative assessment strategies would have no chance of surviving the level of individual motivations to look good and influence local results. The level of opportunism would increase dramatically. Despite the security involved in the PLE, the common frames for essays divulge a great deal of collusion and cooperation for local advantage. Reports from pupils and teachers frequently include comments about the unfairness of buying examinations, cheating, and revision with differential information about exam content.

Pupils steal. The headmasters buy the exam from UNEB. Some bought them [last year].

It is at times unfair on some children. Children who could have been performing well could fail the PLE. The PLE encourages cheating because people want money, so when given, cheating comes about.

Some have cheated. There is malpractice in urban areas. The exams that are set might not be appropriate to our district. They might be set for national interest only.

Sometimes when we receive the envelopes with the scripts, the scripts are fewer than the number of the P7 candidates, possibly because they were swindled or stolen somewhere between UNEB and our school.

Different schools have different problems and yet the PLE does not seem to cater for them. Because the scripts pass through different hands, those in administration sometimes cheat.

National examinations are intended to provide objective indicators of general standards. Opportunism would be high for alternative methods particularly if they were scored by teachers. The stakes are too high for these children in Uganda. Teachers don't want to see their pupils lose out on further education, and they don't want to damage their own reputations. National examination programs offset much of the potential opportunism that would cripple other accreditation approaches. People ultimately still have faith in the outcomes from secure national examinations. No doubt some malpractices occur but the program retains its credibility.

Teachers have *asset specificity*, resulting from their differential training, abilities, and experience. These are particular assets or resources that have value in the context, in this case, education or the school. Experienced teachers may have developed many techniques to help children prepare for traditional examinations. Even those with less experience and more limited repertoires have experienced the traditional examinations successfully and feel more comfortable preparing strategies and methods around the known. New examinations, which teachers have not experienced firsthand and have probably not directly developed, pose considerable demands on teachers. Their craft and experience is devalued in this reform, so they may be reluctant to see the "new" examinations take hold. However, this does not seem to be the case in Uganda. Examinations come from outside the local domains of teachers, and they tend to treat them with great respect and importance. In this context, none of the teachers interviewed wanted to abandon the PLE because of the reform. A few wanted to abandon it because the reform was too slow or too little (e.g., preferring continuous assessment). Also, teachers who knew about the actual change indicated that they were changing their classrooms. They may be aware that their knowledge of the examination increases their skill assets with respect to preparing their own pupils. They may view their early knowledge as advantage enough and value these new assets. But this is a transient advantage, or should be. At the moment, it does matter in Uganda who the teacher is. Schooling with limited or differential regional resources has the inherent problem of unique or imperfectly standardized programs, teachers, and conditions.

We need to also look at the pupil's transactions with respect to the UNEB reform. The pupil in the Ugandan system has considerable bounded rationality in terms of the PLE. There is little information available to pupils about the specific changes that are occurring. The information provided is for teachers. Pupils are therefore dependent upon the teachers for useful examination information. Even if pupils could prepare themselves, they would be bounded by the limits of available information to pupils in general. Where information is readily available, like urban centers, there may be more opportunity for relevant individual initiatives. The question, then, would shift to, could the pupils know what to do to address these changes? The answer is probably not; no more than the teachers do.

Pupils, of course, have strong self-interests in this single-session, high-stakes, national examination series. They would welcome any form of assistance, whether honest or otherwise. Not all would behave opportunistically but because some would, others would be at a

disadvantage if safeguards were not adequate. The opportunism of pupils is likely to be passive, except in cases of direct cheating from another's paper. Because the PLE is short-answer, direct cheating is more difficult than with, say, multiple-choice questions. Like anyone else, pupils are better prepared if they know what the future holds for them. Those pupils who understand the intentions have more of a chance to succeed, or in other language, to exploit the advantage, in these competitive examinations.

In terms of asset specificity, urban children likely have distinct advantages over rural children because of family assets and family education experience. They may have more direct involvement from parents and have more general access to resources that can take them beyond the schooling experience. These assets are not left behind. They use them in the classroom and in examinations. In Uganda, there are considerable differences in assets. This is a difficult difference to reconcile in a national examination program that is supported so minimally in terms of relevant teacher training and pupil assistance. It matters who the pupils are and what family resources accompany them.⁴¹

One could say that the Ugandan reform was doomed to no better than modest success. For some, this modest success may serve them well in their futures and in contributions to society. But for most teachers and pupils, there is an element of unfairness in this reform, derived unconsciously from differential information and preparation, i.e., different opportunities to learn, that remains worrisome. Many pupils forfeit their chances through sickness, fear, stress, apathy, lack of preparation, or many other reasons attributable to systemic elements of the competitive program and not necessarily related to inadequacies of talent. Traditional testing on knowledge is well known, seems accessible to all, and has a credible base in the symbolic meaning of schooling for most folks. For those who want more from our schools, this is a powerful constraint. Examination reform turns out not to be an OLB. It has many dimensions and many bumps.

The idea that has shaped modern American schooling is to reduce the stress load on pupils through continuous assessment using alternative or authentic assessment approaches. More observations of performance also can improve reliability and comprehensibility of the assessments. The bounded rationality accompanying each assessment is less than on a large-scale national exam, but there is increased opportunism if there is a high-stakes conclusion. The relationships between teachers, pupils, and communities in Uganda would make it difficult for teachers to form or enact an objective decision in light of the many counter incentives. A solution now under trial in Vermont is to couple alternative assessments with standardized school examinations through sampling to reduce opportunism. Performance assessment by teachers poses considerable demands on teachers, who in this context have limited skills and experience with this kind of assessment. Considerable staff development would need to accompany such a

⁴¹Asset advantage in a competitive economy was one of the early reasons educational expansion was widely supported. Principles of equality will actually slow the expansion because they work against enhanced individual interests and resulting personal advantages due to stratification and special incentives. See Meyer, J. (1992). The social construction of motives for educational expansion. In B. Fuller and R. Rubinson, *The political construction of education: The state, school expansion, and economic change*. NY: Praeger, 225-238.

reform, and this is expensive and long-term. Work in Swaziland using continuous assessment imbedded within instructional modules has demonstrated the possibility of success (project with the Institute for International Research). The assessments are developed in standardized form for everyone using the materials. They are not for high-stakes decisions, however. Performances are used in self-and classroom-assessments for enrichment, remediation, and brightening. In this way, the low-stakes feedback from continuous assessment enhanced scores on standardized criterion-referenced tests.

Asset equality is perhaps impossible to accomplish in practice. Individuals will always come with differential resources. It will matter who the teacher is and who the pupil is. But some of these differences can be addressed by attention to the opportunity to learn, as now being formulated in American education. Currently, the *Basic Education Support Project* in Namibia is geared toward the preparation of teachers and pupils in disadvantaged schools to offset differential base knowledge between these rural schools and others across the country. The idea is to reduce asset differences. Only then do these pupils have a chance in the national competition for places.

There is little doubt that examinations are essential components in the reform process. Simplistic notions of reform, such as changing the perspective of exams, can produce substantial inequities in opportunities dependent on exam performance or moving to alternative assessments that abandon safeguards to opportunism and depend upon the judgments of teachers with considerable bounded rationality in this new context can create credibility problems for system efficacy. In addition, examinations fall prey to what Snodgrass and Biggs⁴² have called the *double development squeeze*, which in this case refers to the Ministry's dilemma that schooling costs are rising significantly but jobs do not exist for large numbers of graduates and thus the value of increased access declines. Technological diffusion increases the complexity of jobs and requires higher job entry skills and talents. Because of increasing teacher salaries and rising populations, the cost of schooling is also increasing at the lower levels of schooling, where staffing and population are the greatest. Ministries of education need to make difficult choices about where to invest. "Education for all" limits the decision-makers' options. Primary levels are growing so rapidly that few funds exist for policy considerations at other levels (e.g., Namibia schooling uses over 30 percent of the national budget). Examinations help decide on the pupils in whom the system can invest. The tragedy is that, for many countries, people are the key resource and limiting access to higher levels of quality education curtails national options and economic growth. It is not clear where the value of education lies for society, or if such a question can even be answered in its general form. The notion of reform in schooling and education is inherently complex. Examinations address some of the issues within a particular context and time, with mixed success, but the questions do not go away.

Two dimensions that House did not consider from Williamson were *uncertainty* and *number*. *Uncertainty* refers to adaptations to changes over time and condition. UNEB is never

⁴²Snodgrass, D.R., and Biggs, T. (1996). *Industrialization and the small firm*. San Francisco, CA: International Center for Economic Growth and Harvard Institute for International Development, 70.

sure how teachers and pupils will react and what specific adaptations they will make to examination reform, and correspondingly, the teachers and certainly the pupils have very little information about UNEB's adjustments in the exams or exam content. Williamson (1985) cites Oskar Morgenstern's (1976) Holmes-Moriarity dilemma:⁴³

Sherlock Holmes, pursued by his opponent, Moriarity, leaves London for Dover. The train stops at a station on the way, and he alights there rather than traveling on to Dover. He has seen Moriarity at the railway station, recognizes that he is very clever and expects that Moriarity will take a faster special train in order to catch him in Dover. Holmes' anticipation turns out to be correct. But what if Moriarity had been still more clever, had estimated Holmes' mental abilities better and had foreseen his actions accordingly? Then, obviously, he would have traveled to the intermediate station. Holmes, again, would have had to calculate that, and he himself would have decided to go on to Dover. Whereupon, Moriarity would again have "reacted" differently. Because of so much thinking they might not have been able to act at all or the intellectually weaker of the two would have surrendered to the other in Victoria Station, since the whole flight would have become unnecessary.

Uncertainty in the Uganda reform could be reduced by better diffusion tactics (e.g., Rodgers, 1995).⁴⁴ UNEB made some attempt. Limited resources permitted only three publications of the newsletter over five years, so penetration was incomplete. The newsletters were mentioned by some of the teachers, and many thought it a good way to communicate, if regular publications were possible.

Number refers to the number of transactions, and in this case, the likely success of passing the national examinations and selection for the next level. For the PLE, there is only one opportunity to take the examinations, and few get through the exams to the next level. This really defines the notion of *high-stakes*. The value of the stakes plays an important role in reactions and impact. Low-stakes options may carry more educational value, while high-stakes carry high personal value. For Uganda, the consequences of the PLE are considerable for pupils and teachers.

Strange Loops, Complexity, and Institutional Credibility

In education reform, the same problems seem to arise each time we look carefully at the area or context. The seriousness or priority of particular items may rise or fall from time to time but no item can be completely ignored. Problems return! Even as we "solve" them, they arise again, or seemingly so. The apparent circularity is a result of working within the system with

⁴³Morgenstern, O. (1976). Perfect foresight and economic equilibrium. In S. Schotter (Ed.), *Selected economic writings of Oskar Morgenstern*. NY: NYU Press, 173-174.

⁴⁴Rodgers, E.M. (1995). *Diffusion of innovations*. 4th ed. NY: The Free Press.

reduced problem dimensions (because of bounded rationality) to manage identifiable, well-structured problems. This is called a *strange loop*.⁴⁵

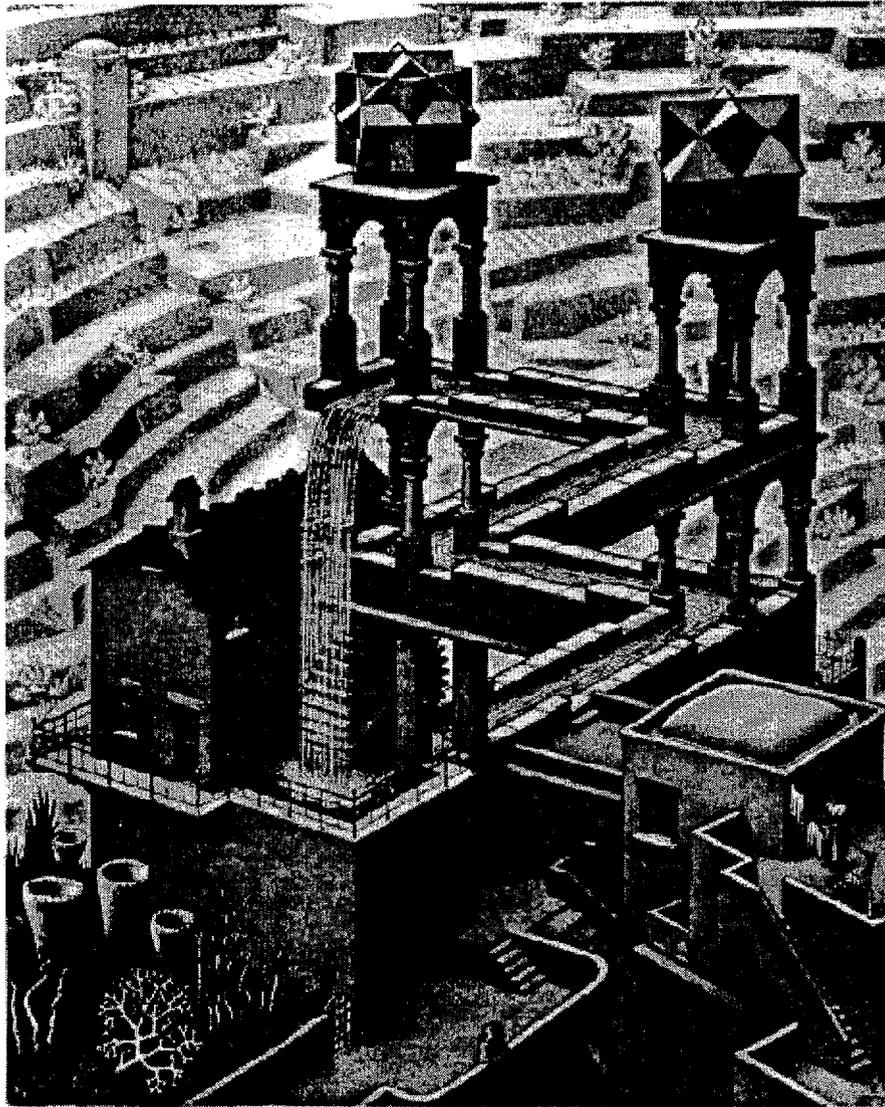


Figure 4. *Waterfall*, by M.C. Escher, 1961, illustrating a “strange loop” with a reduction in dimensionality.

Escher’s *Waterfall* is a six-step endlessly falling loop, such that the water “begins” in the pool at the side of the house, flows down four canals, and falls back into the pool to flow again down the canal. The illusion was discovered by Roger Penrose in 1958, and results from the reduction of the event to a depiction in less dimensions or the tangled hierarchy where high

⁴⁵Hofstadter, D.R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. London: Penguin Books.

levels reach down to a lower level. Coupled with the illusion is the notion of infinity, so that this presentation allows the representation of an endless process in a finite space (see Hofstadter, 1979, for the development of these ideas and some wonderful applications of this approach). We have the aphorism attributable to Alphonse Karr (1809-1890) that “the more things change, the more they stay the same.” When we link education problems with these notions, we find that the solving of problems in education systems is infinite, a never-ending story, and the “solutions” change the context but similar macro problems return to continue forever to demand our attention.

The strange loop is a product of our own bounded intellect and the tangled hierarchy of the elements of an education system. The tangle comes from the fact that there’s no inviolate level to which to appeal the validity of a reform. Teacher educators comment on curriculum; curriculum developers comment on examinations; examinations reflect on teachers; teachers assess pupils; and so on. There is no final appeal for ultimate judgment. The choice is always ambiguous. We cannot deal completely with the complexities inherent in education systems, so we reduce their complexity and artificially structure the problems so that we can handle them. John Dewey characterized it as the Western culture’s “quest for certainty.” Sometimes, in fact, we deal with them as simple, independent problems with answers. Problem solving strategies can wreak havoc. Particular solutions exacerbate other problems. This is a perennial problem in development projects, for example. They are always planned as rational agendas for reform, with little emergent planning possible or encouraged (although sometimes forced). The emphasis is on deliverables, which presuppose solutions.

Education problems are complexly organized, ensnarled webs of problems that are ill-structured.⁴⁶ Complexly organized problems are⁴⁷:

- ◆ highly interdependent, such that exams are linked to curriculum, teacher education, and so on;
- ◆ complicated, such that exams themselves have many aspects (e.g., form and content) to them with relationships among their components;
- ◆ dynamic, such that exams are part of an ever changing system;
- ◆ ambiguous, such that there is no single correct view about exams or their relationship to the other elements of the system;
- ◆ political, such that there will always be competing claims about education and about exams; and

⁴⁶Mason, R.O., and Mitroff, I.I. (1981). *Challenging strategic planning assumptions: Theory, cases, and techniques*. NY: John Wiley and Sons.

⁴⁷Rirrel, H. (1972). On the planning crisis: Systems analysis of the ‘first and second generations.’ *Bedriftsokonomien*, Number 8, 390-396. (Cited in Mason and Mitroff, 1981).

- ◆ limited by system constraints, such that the capabilities and bounded rationality of the local context must be taken into account when addressing the role and operationalization of exams in the program.

Examinations are part of the systemic complexities—exams are interdependent with other aspects of education; reform of examinations changes the context and requires changes in other elements of education which in turn change the context. Ultimate solutions are not possible. Complexly organized problem sets require management that is on-going, strategic, and reflective, with continual attention to the improvement of the system's information architecture and increased capacity through training to manage, lead, and understand the education system. Their complications and interrelatedness must be understood, making research and reflection important features in dealing with them. Their dynamic character means that risk is essential because they will not respond to static rationalism. Because they are inherently ambiguous, they invite multiple perspectives and require maximum consultation. Their political tradeoffs minimize the likelihood of fully cooperative arrangements. And the system's constraints limit the responsiveness to the problem set. As summarized by Hofstadter (1979, p. 693): "...the ultimate explanation of a society's political behavior always lies at the *grass roots level*." The management of education is "...a polite gesture granted collectively by millions of people—and it can be overridden just as easily as a river can overflow its banks."

Learning does not require examinations. But it seems that schooling systems do. Examinations become the defining entities for an education system. Institutional beliefs that establish the credibility for the system must be considered in any reform design. This includes the cultural and symbolic aspects of the system. One of the problems in understanding education reform is that institutional success is not always congruent with technical progress. Examinations legitimate the educational enterprise in Uganda, and Ugandan schools and teachers have survived enormous stresses and incursions in the cause of education. Uganda has a long tradition of education that is often reflected in individual successes from those passing the "high standards" examinations. But the examinations remain remote visions to the teachers of Uganda. Individual advantages tend to go to those who achieve well throughout their educational careers and may have less to do with formal instructional programming. At the higher cognitive levels, there is no particular linkage between what an examination asks and what a teacher necessarily does. As the testing literature documents, specific changes in an examination, especially in recall items, can emerge in classroom coverage. But complex cognitive processes demand different cultural and dynamic features in classrooms that are unlikely to emerge spontaneously. It's literally a change in perspective. Examinations are essential for reform, but not sufficient. They mark the legitimization of the system but not necessarily the validity of its technical competence.

Advocates for alternative educational assessment recognize these deficiencies in high-stakes, single-occasion, narrowly derived measures of performance and understanding. But they fail to fully recognize the role of national examinations in the survival and "success" of emerging central education systems. Examinations are part of the ritual of schooling. Potentially dramatic individual dysfunctions are tolerated because of the mobilizing and accrediting aspects of examinations to substantial parts of the internal polity and the universal educational community.

Organizational systems are highly dependent on their political and cultural environments and on the support and legitimacy that they are accorded in those environments (Meyer and Scott, 1983).⁴⁸ Examinations “guarantee” that legitimacy, regardless of their connection to actual technical aspects of the system. Emerging systems, in particular, are likely to exhibit considerable fragmentation and inconsistency among their organizational components designed to implement the educational mission. This “loose coupling” can be an advantage where many elements are deficient and resources for many activities inadequate. National examinations “permit” the continuation of many different and uncoordinated efforts to continue without system disintegration. The Ugandan education system has survived nearly unimaginable difficulties over the last several decades and examinations have defined the system in the absence of many, usually considered important, features and resources.

Reliance on national examinations for credibility provides a shorthand for harder-to-attain technical competence. Whatever the exams say is what characterizes the local environment in achievement. All other assessments accrue over time (e.g., success of graduates, national accomplishments, societal stability, human rights records, individual achievements, schooling reputation). Countries tend to risk some credibility at the lower levels of the system, but tie upper level examinations to international syndicates. The associated credibility buffers the entire schooling system from concerns with inefficiencies, inconsistencies, and limited technical competence. Without the linkages, the system would have to argue for credibility on its own internal merits, and this has not been without its difficulties even for highly developed systems (like that in the United States, where credibility problems abound). Changing perspective to *thinking* instructional programs alters the nature of the system and examinations, and threatens to expose the system’s inadequacies and damage credibility. Uganda proceeded slowly, but even so, UNEB has become embroiled in political intrigue which threatens the existence or investment in the reform. Other political factors have loomed large in UNEB’s demise but the charges of excessive opportunism have damaged the influence of the reform with schools. Why change when advantage is gained unfairly (or perceived to be so), and change is so difficult? Even those teachers who understand the nature of examination reform do not or cannot change their instructional programs to meet the new needs of the reform. They make minor concessions or adjustments to their approach. The dynamics of classroom instruction, meanwhile, remain the same.

As suggested by Hofstadter’s reference to grass roots as the level for external validation of an education strategy (or any political strategy), the possibilities for reform are limited by the tolerance and bounded rationality of the society supporting the education system. In strong government contexts, where there are high levels of internal organization, it may be possible to ignore the populace and deal directly within the schooling system. Otherwise, and for long-term success, the grass roots should be partners in the reform. Citizens tend to be conservative in choices on reform, and we have seen that education’s complexly organized system of problems

⁴⁸Meyer, J., and Scott, W.R. (1983). *Organizational environments: Ritual and rationality*. Beverly Hills, CA: Sage.

do not lend themselves to simple, problem-solving solutions. Risk is necessary. Betty Steffy points out, from the experience in the Kentucky reform:⁴⁹

It is a fallacy that comprehensive educational reform can ever be crafted by people unschooled about schooling, ignorant of educational history and methodology, unable or unwilling to deal with how the future can be shaped into schools instead of using the past as the key to reshape schools. The past is dead. The future is not clear. The people cannot be expected to know how to translate uncertainty into an institution, craft its ethos, formulate methodologies that run true to that ethos, evaluate successes and failures, and engage in constructive alterations when things don't exactly work as planned. Providing the education required to produce bottom-up school restructuring will require more time and training than is currently available to most parents and teachers.

The key is gaining public support so that changes in the instructional program will have a chance to survive the politics of education. Robert Rothman ends his treatise on *standards, assessment, and school reform* in his book *Measuring Up* (1995, pp. 186-187):

To generate public support, educators must make the case that the current expectations for student performance are inadequate. They must emphasize that a functioning democracy and an increasingly technological, high-skills economy demand graduates who can reason, solve problems, and communicate. They must show that we have learned that such skills are fundamental to learning. And they must explain that developing those abilities, and determining whether students have attained them, requires new methods of assessing student performance. ...The road to the twenty-first century begins with a new way of knowing what young people know.

Many developing countries have centralized systems, including Uganda. One might expect that there would be tighter internal organization and autocratic control, and less need to move outside the system for support. This proves incorrect.⁵⁰ In the end, teachers and communities decide what happens in the local schools. Without their encouragement and support, changes in perspective will not appear in classroom dynamics. Even with support, there is a long way to go to enable the system to respond appropriately to examination reform. For Uganda, the attempt at examination reform, in its simplest form, was worth the try. The Uganda reform was the special project of a few talented and insightful people at UNEB. They exploited the ambiguities of political uncertainty to carry out the reform. As the political machine developed, these individuals had little political base in the region where UNEB resides. The reform attracted little attention in this case but the reformers were targets of political power restructuring. The future of the reform is unknown.

⁴⁹Steffy, B. E. (1993). *The Kentucky education reform: Lessons for America*. Lancaster, PA: Technomic Publishing Company, 265.

⁵⁰For example, Meyer, J. (1995). *Organizational integration in Lesotho Primary Education: Loose coupling as problem and solution*. Maseru, Lesotho: Ministry of Education and Primary Education Project. Available from John Meyer at Stanford University.

Perhaps in time, with attitudes changed among a large cadre of teachers, skills and resources will follow and effect the reform. Like the butterfly flutter that affects weather systems, this effort has changed something about the education scene in Uganda forever. A small change should not be underestimated for the future. But the complexities cannot be overlooked either.

Conclusions

Innovations sometimes seem to work locally or in parts (like each canal in the Escher *Waterfall*), only to leave a trail of a strange loop when carefully examined overall. Ugandan examination reform, when treated like an OLB with inadequate considerations of education's complexities, affects pupils in many unintended ways, has limited impact on improved classroom instruction, and risks the credibility of the system. We end up with the same web of problems unaddressed. Table 7 presents our conclusions on the dimensions outlined in this paper. In the presence of these conditions, the set of solutions suggested by transaction cost economics calls for increased governance with high levels of internal organization. For many reasons, both technological and historical, education systems remain loosely coupled organizations that are highly resistant to tightening strategies of internal organization. This strategy has to be translated within the education context. The study of what to tighten and what to loosen is nontrivial.

Uganda cannot legislate or administer increased actions in instruction that are compatible with higher level cognition instruction to follow the changes on the examinations, and they would have no way to afford the vast resources necessary for accountability or supervision in such an organization. They have, in fact, done more than most countries would attempt. This was possible because UNEB had little political interference at the time. High governance would increase the likelihood of diffusion but the change-in-perspective suggested might prove too risky in terms of system credibility. Well-being and stability in education, a highly visible and intrusive central government function in Uganda, is associated with the well-being of the state and the quality of governance. Sudden drops in exam performances and changes in school and teacher status would introduce many surprising instabilities. Uganda had many other problems to occupy the populace so the UNEB reform proceeded without opposition. When we talk to the teachers and headteachers of Uganda, many know what is being asked for in the reform. They even say that they have changed accordingly, although we cannot substantiate that. Education is taken very seriously in Ugandan communities. With the will of most teachers to teach well and help pupils succeed in exams and life, there is the beginning of a critical mass of adopters to set the reform in motion. As Rodgers points out for the now ubiquitous facsimile adoption:

...Fax was invented in 1843 by Alexander Bain, a Scottish clockmaker who called it a recording telegraph because the message was transmitted over telegraph lines. There were no adopters....Although the fax boom began in the United States around 1983, the rate of adoption remained quite slow until 1987, the year in which a critical mass of users occurred....Starting in 1987, Americans began to assume that 'everybody else' had a fax machine....So it took [about] 150 years for fax to become an overnight success!

Nearly all of our interviewees profess awareness-knowledge of the reform. The rate of adoption and implementation usually lags behind. We do not have a national sample so we cannot

estimate the system rate. For our group, this is a large number of adopters. If eventually, the reform takes hold in action in the classroom, the questions are, “will it be supported or allowed to remain,” and “what are the implications for other aspects of the system.”

A long-term answer for systemic change lies in the polity (the grassroots). If the polity understands the issues of perspective and will tolerate or encourage the restructuring of the education system (perhaps where the political costs are assigned to a past regime), then top-down reform through examinations may work. Eventually, even in the Ugandan situation, top officials, outside of UNEB, need to adopt the reform. Issues of bounded rationality, opportunism, asset specificity, uncertainty, low numbers, and institutional culture, of course, must be attended to. Chapman, Mählek, and Smulders⁵¹ remind us that national examinations are one of the few reform measures available for central government to leverage change at the classroom level. This study of the Uganda examinations reform reminds us that change is difficult to control, and education reform is complex, subject to strange loops, and frequently reinterpreted in light of institutional meanings and rituals. No OLBs need apply.

⁵¹Chapman, D.W., Mählek, L.O., and Smulders, A. (1997). *From planning to action: Government initiatives for improving school level practice*. Paris: International Institute for Educational Planning.

Tables

Table 1. Characteristics of Test Format

True/False	Multiple-Choice	Short-Answer	Essay	Project	Performance Based	Dialogue-Oral
<i>Less realistic</i>						<i>More authentic</i>
Does the testing task resemble an actual application of the learned material? Multiple-choice and true/false items can be rather artificial.						
<i>Lower-order cognitive skills</i>						<i>Higher-order cognitive skills</i>
The lower-order cognitive skills deal with largely factual material. Higher-order cognitive skills include being able to understand, analyze, synthesize, evaluate, and use information.						
<i>Objective scoring</i>						<i>Subjective scoring</i>
Subject-matter experts would largely agree on a pupil's score for most multiple-choice or true/false tests. There would tend to be much less agreement among the same experts if the test were an essay or performance assessment; in fact, there may be more than one correct answer to a question.						
<i>Requires recognition</i>						<i>Requires Recall</i>
For selected response items, the examinee must only recognize the correct answer rather than produce it.						
<i>Smaller learning component</i>						<i>Larger learning component</i>
There is little (if any) learning that takes place in a true/false test, but writing a paper or putting together a project are commonly used instructional devices.						
<i>Less diagnostic (summative)</i>						<i>More diagnostic (formative)</i>
Incorrect responses to a multiple-choice item may give some indication of a learning difficulty, but constructed responses are usually more informative in this sense.						
<i>More reliable</i>						<i>Less reliable</i>
Although all formats are capable of being more or less reliable, the added consistency of rater disagreement tends to make those formats with subjective scoring less reliable.						
<i>More content coverage</i>						<i>Less content coverage</i>
In general, when available testing time is used for greater depth (e.g., the essay format), there will be fewer items and less breadth of coverage. When individual items require less time to complete (e.g., true/false format), it is possible to use more items within the time limit to cover a larger span of the content domain(s).						
<i>Guessing more of a factor</i>						<i>Guessing less of a factor</i>
It is easier to guess correctly on recognition items and more difficult to guess correctly on those where the correct answer must be recalled or produced. On the other hand, faking can be easier on essays than it is on a multiple-choice test.						
<i>Cost in the beginning for test construction</i>						<i>Cost at the end for test scoring</i>
It takes more time to evaluate items with subjective methods, but objectively scored items may take more time to construct.						

<i>Focus on the product</i>						<i>Focus on the process</i>
-----------------------------	--	--	--	--	--	-----------------------------

This difference is especially evident with performance-related measures.

<i>More emphasis on behavioral learning theories</i>						<i>More emphasis on cognitive learning theories</i>
--	--	--	--	--	--	---

Behavioral learning theories emphasize identifiable elements that lend themselves to specification for inclusion in true/false and multiple-choice tests. Cognitive learning theories do not necessarily.

<i>Absolute standards: Mastery is usually possible</i>						<i>Relative standards: Mastery may not be possible</i>
--	--	--	--	--	--	--

When there is a clear, well-defined, and finite content domain, the results can be reported in terms of percent mastery of the domain (absolute standard). When the content domain is unlimited and/or difficult to define, referencing individual performance to the performance of a group will be more meaningful (relative standard).

<i>More instructional sensitivity</i>						<i>Less instructional sensitivity</i>
---------------------------------------	--	--	--	--	--	---------------------------------------

Factual tests tend to be more instructionally sensitive. That is, gain scores are typically greatest when the objectives of an instructional program are largely knowledge-based.

<i>Individual differences may be created by the test</i>						<i>Individual differences are assumed real</i>
--	--	--	--	--	--	--

Either individuals tend to differ in level of cognitive skills or the tests (usually norm-referenced) are constructed from items which exaggerate (possibly educationally unimportant) differences.

<i>Strict alignment of teaching and testing</i>						<i>Novel problems or applications on the test</i>
---	--	--	--	--	--	---

Objective test formats seem to lend themselves to instructional designs that align instructional objectives, in detail, with testing objectives through precise item specification procedures. Essay and oral tests lend themselves to those cases where we wish to explore the application of principles to new or unseen problems.

<i>Suited to small or large-scale applications</i>						<i>Suited to small-scale applications</i>
--	--	--	--	--	--	---

The cost per pupil, reliability, and summative orientation of the objectively scored item are factors in this.

<i>Instruction decreases pupil differences</i>						<i>Instruction increases pupil differences</i>
--	--	--	--	--	--	--

With a finite domain, instruction can decrease pupil differences as mapped by objectively scored items. As the profiles change the item selection also changes to keep those items in place that maximize the appearance of individual differences. When the domain is unlimited, the influence of instruction on individual pupils varies substantially and individual differences increase.

<i>Drill and practice are the rule</i>						<i>Discovery methods of learning are preferred</i>
--	--	--	--	--	--	--

Where there is much information to be gained, the *process* takes second seat to the *product*. This is quite difference from the view that learning is a process.

<i>Less likely to be used for ability grouping</i>						<i>More likely to be used for ability grouping</i>
--	--	--	--	--	--	--

Ability grouping focuses more on process skills (special education, gifted, in class *bluebirds*).

<i>Teacher provides structure</i>						<i>Pupil provides structure</i>
-----------------------------------	--	--	--	--	--	---------------------------------

Many tasks require the pupil to bring structure or organization to bear on a problem. This is often part of what is being evaluated. Compare this to the highly (teacher) structured objective test.

<i>Pupil is often passive (cognitively)</i>						<i>Pupil is more active (cognitively)</i>
---	--	--	--	--	--	---

Greater demands are placed upon the examinee who faces an open-ended test format. There is greater latitude to the way the problem is solved and possibly the number of correct responses. Test scoring becomes difficult because examinee creativity (or bluff) is more likely.

<i>More grade retention</i>						<i>Less grade retention</i>
-----------------------------	--	--	--	--	--	-----------------------------

The tendency to retain in grade those pupils who have not mastered the minimum content is greater when those skills are seen as necessary for the effectiveness of the next level of instruction.

<i>Less test complexity</i>						<i>Greater test complexity</i>
-----------------------------	--	--	--	--	--	--------------------------------

Essays and oral dialogue provide greater opportunities for many different topics and relationships to emerge in the answers provided. Related to less structure, these formats can explore the interrelationships among domains and their application in a variety of contexts.⁵²

⁵²This table is based on work presented to the American Educational Research Association by George Johanson, 1996.

Table 2. Tests in Recent Reforms in the United States

<i>Context</i>	<i>Program Features</i>	<i>Reactions/Evaluations</i>
<p>Colorado (1989)</p> <p>Mark Twain Elementary,</p> <p>Littleton High School, and Heritage High School</p>	<p>Standards calling for the demonstration of the ability to reason, solve problems, and communicate, and develop new assessments that included having pupils produce questions, cooperative and collaborative discussion and data collection, working independently and collectively, and writing displaying their own reasoning. Assessments apparently influence teaching—teachers seldom rely on textbooks; open standards; new report cards with 95 categories and more information.</p>	<p>Parental backlash against standards and assessments for reform. Standards focused on thinking but parents wanted knowledge of core information. Charged assessments appeared less rigorous and relied heavily on subjective teachers' judgments.</p> <p>Despite victories on school boards, not all reforms were 'thrown out.' Prohibited assessments in areas of attitudes and values.</p>
<p>California (1990)</p> <p>O'Farrell Community School</p>	<p>Portfolios; restructured school; reduced class size by doing away with assistant principal and counselor. Located in area of gangs and drugs. The 'O'Farrell Standard' consisted of a set of tasks that all pupils had to complete to graduate. Created a 'portfolio culture.' Teachers explicitly link tasks to relevance later.</p>	<p>Parents didn't complete their evaluations of children's work so this part was dropped. Attendance rate has improved. Teachers want to teach there.</p>
<p>Vermont (1990)</p> <p>Planned for forty-eight school districts; another ninety asked to participate themselves</p>	<p>Portfolio assessment with teacher support. Extensive program of portfolios that took a considerable amount of time to implement.</p> <p>Program was voluntary.</p>	<p>Educators considered the assessments a worthwhile burden. Teachers and principals reported the program changed instruction.</p> <p>Evaluation by RAND. Positive except that assessment scoring was found wanting. Portfolios were effective instructional tools but not good indicators of pupil performance.</p> <p>Public concern prompted return of standardized tests. Involuntary standardized tests introduced to check on alternative assessment strategy.</p>

<i>Context</i>	<i>Program Features</i>	<i>Reactions/Evaluations</i>
<p>Kentucky (1990)</p> <p>Approximately 1,350 schools across the state</p>	<p>Kentucky Education Reform Act (KERA) restructuring the entire system.</p> <p>Included assessment of achievement goals; development of statewide assessment program; publication of annual performance report by districts.</p> <p>Interim assessment system administered in 1992: writing portfolio, performance event, and transition test. Now NAEP performance tasks, multiple-choice and short-answer tests, and portfolios.</p>	<p>Districts now faced with aligning curriculum with state assessment. Seventy-five value outcomes related to each of six learning goals (not based on subject areas). Some improvement at 4th and 5th grades but decline at 12th.</p> <p>Before this, curriculum driven by Kentucky Essential Skills Test, California Test of Basic Skills, textbooks, and Kentucky Program of Studies. Commitment apparent for new assessment methods.</p>
<p>Pittsburgh (1987)</p> <p>Harvard and ETS created Arts PROPEL</p>	<p>Used portfolios to encourage pupils to reflect on work and monitored pupil performance over time.</p>	<p>Pupil work improved as they learned the standards of performance. Portfolios, as performance measures, were problematic technically. Public confusion over portfolios also resulted. Used external markers to gain seal of approval.</p>
<p>California (1987)</p> <p>California Assessment Program became California Learning Assessment System in 1991</p>	<p>CAP featured writing assessment using essays and open-ended questions in mathematics.</p> <p>Created curriculum frameworks in core subjects to outline the program of instruction. CLAS placed substantial reliance on new forms of assessment, including on-demand portfolios. Contracted with CTB, Riverside, Psych Corp, and ETS.</p>	<p>Very low results initially triggered criticisms of the types of knowledge and skills in the new assessments, both in terms of particular content and in terms of the breadth of assessment, which covers attitudes and beliefs.</p>

<i>Context</i>	<i>Program Features</i>	<i>Reactions/Evaluations</i>
<p>Pennsylvania (1992)</p> <p>Outcomes-Based Education</p>	<p>Replaced requirements for course credits with learner outcomes that had to be demonstrated. Defined what pupils should know or be able to do in 55 outcomes with goals in 9 academic areas. These were less detailed than the California frameworks.</p>	<p>Unions questioned role of local school boards, which might threaten programs and jobs.</p> <p>Critics concerned with measurement of attitudes and beliefs, and requests for background information. Privacy concerns, and teacher credibility questions.</p> <p>Outcomes-based education became law in 1993. OBE is now politically charged because of resistance to the new standards and assessment methods. Worry about job-market reaction, rigor of reporting, reliance on teachers' judgments, and impact on minorities where objectivity has been compromised.</p>

Sources: Steffy (1993) and Rothman (1995).

Table 3. Rubrics for Observation Variables

C = clarity TP = thoughtful practice M = motivation F = feedback

<i>Rubrics for the Focal Variables in Classroom Observations</i>				
Variable	Ratings (Low ⇒ High)			
C1 Lesson Objectives Objectives of lesson are clear and well understood.	No Objectives.	Objectives implicit.	Objectives present.	Objectives are highly clear.
C2 Appropriateness of Activities Activities support lesson objectives.	No supportive activities.	A few supportive activities.	Some supportive activities.	Most or all activities support objectives.
C3 Elaboration and Examples Elaboration and good examples are used throughout lesson.	No or poor elaboration or examples.	Minimal elaboration and examples; may be weak.	Moderate use of elaboration and examples; may be good.	High quality examples and elaboration used throughout lesson.
C4 Knowledge of Content Knowledge is accurate and thorough.	No substantive knowledge.	Minimally accurate and thorough knowledge.	Moderately accurate and thorough knowledge.	Highly accurate and thorough knowledge.
C5 Coherence and Content Flow Content elements of lesson fit together without disjuncture. Lesson contents form a coherent package.	No coherence. Elements are independent pieces.	Minimal coherence. Elements fit but with considerable disjuncture.	Moderate coherence. Elements fit adequately together.	High coherence. Elements fit very well into total package.
C6 Process and Events Flow Continuity of process. Events of lesson form smooth sequence.	No flow to process.	Minimal flow to the process.	Moderate flow to the process.	High level of flow to the process.
TP1 Opportunities for Practice Multiple opportunities for practice by each pupil.	No opportunities for any pupil to practice.	Minimal opportunities for each pupil to practice.	Some opportunities for each pupil to practice.	Multiple and frequent opportunities for each pupil to practice.

C = clarity TP = thoughtful practice M = motivation F = feedback

<i>Rubrics for the Focal Variables in Classroom Observations</i>				
Variable	Ratings (Low ⇒ High)			
TP2 Opportunities for Thoughtfulness Multiple opportunities for thoughtfulness by each pupil.	No opportunities for any pupil for thoughtfulness.	Minimal opportunities for each pupil for thoughtfulness.	Some opportunities for each pupil for thoughtfulness.	Multiple and frequent opportunities for each pupil for thoughtfulness.
TP3 Cognitive Level of Learning Lesson emphasizes critical analysis.	Memorization of facts.	Understanding of facts.	Active use of knowledge.	Critical analysis of facts and knowledge.
TP4 Teacher Preparedness and Organization Teacher is well prepared. Functional elements are arranged for effective learning (teacher, pupils, work stations, materials, aids).	Teacher not prepared. Elements not arranged. Much time wasted.	Weak teacher preparation and arrangement of elements. Some wasted time.	Some preparation is evident. Elements arranged adequately. Little time wasted.	Teacher well prepared. Elements well arranged. No time wasted.
TP5 Pupil Participation Pupils personally participate in one of more activities (group work, manipulation of materials, etc.).	Pupils are passive and only listen, sit or watch.	A few pupils actively participate while most watch.	A moderate number of pupils actively participate.	All or most pupils actively participate.
TP6 Questions Involving All Pupils Knowledge of all pupils is explored through teacher's questioning of individuals or groups.	Teacher asks no questions.	Teachers asks questions of whole class only.	Teachers asks questions of some individual pupils.	Teacher asks questions of all pupils, individually or in groups.
M1 Quantity of Teaching Strategies A number of different teaching strategies is used to provide opportunities for pupils to become engaged in learning (lecture, dialogue, group work, projects. etc.).	Single teaching strategy that provides no opportunity to become engaged.	Single teaching strategy that provides a few opportunities to become engaged.	Two teaching strategies that provide some opportunity to become engaged.	Multiple teaching strategies that provide considerable opportunities to become engaged (3 or more).

<i>Rubrics for the Focal Variables in Classroom Observations</i>				
Variable	Ratings (Low → High)			
M2 Quality of Teaching Strategies The quality of teaching strategies enables pupils to become engaged in learning (dynamic presentations, questions, inquiry, rehearsing, examples, reinforcing activities, etc.).	Strategies do not enable pupils to become engaged.	Strategies enable minimal pupil engagement.	Strategies enable some pupil engagement.	Strategies enable considerable engagement.
M3 Pupil Attentiveness High percentage of pupils are on task, absorbed, interested, not daydreaming, or playing.	No or almost no pupils are on task. 0 ----- 25	A few pupils are on task. 26 ----- 50	A moderate percentage of pupils are on task. 51 ----- 75	Many or all pupils are on task. 76 ----- 100
M4 Teacher Enthusiasm Teacher communicates interest, enthusiasm, and excitement about subject through voice, pace, nonverbals, etc.	Teacher communicates dislike for subject.	Teacher communicates a little interest in subject.	Teacher communicates moderate interest in subject.	Teacher communicates high interest.
M5 Pupil Enjoyment High level of enjoyment is shown by pupils.	All or almost all pupils seem bored, apathetic, off-task, passive, or unhappy.	Minimal interest or little enjoyment shown by most pupils.	Moderate enjoyment shown by some pupils.	High level of enjoyment shown by all or most pupils.
M6 Pupil Initiative Pupils initiate learning activities and demonstrate creativity.	No initiative or creativity shown.	Minimal initiative or creativity shown by some pupils.	Moderate amount of initiative and creativity shown by some pupils.	High level of initiative and creativity shown by all or most pupils.
M7 Teacher Responsiveness to Pupil Initiative Teachers respond positively to pupil-initiated learning activities.	Teacher ignores or squelches all pupil-initiated learning.	Teacher responds positively to one or two pupil-initiated learning.	Teacher responds positively to three or a few pupil-initiated learning activities.	Teacher encourages pupils to initiate their own activities and responds positively when they do.

C = clarity TP = thoughtful practice M = motivation F = feedback

<i>Rubrics for the Focal Variables in Classroom Observations</i>				
Variable	Ratings (Low → High)			
M8 Complexity and Interrelatedness Information is complex. Different points and perspectives are interrelated.	No complexity or interrelation.	Minimal complexity and interrelation.	Moderate complexity and interrelation.	Considerable complexity and interrelation.
M9 Cognitive Challenge Level of learning positively motivates all or most pupils at appropriate level. Lessons are neither too easy nor too difficult for pupils.	No pupils or only a few are challenged by level of information or tasks.	A small number of pupils are challenged.	A moderate number of pupils are challenged.	Most pupils are challenged.
F1 Quantity of Informative Feedback Teacher uses frequent feedback to give pupils constructive evaluation of their performance.	Acceptance of pupil performance without feedback.	Infrequent informative feedback about performance.	Moderate amount of informative feedback about performance.	Frequent informative feedback about performance.
F2 Quality of Informative Feedback Teacher uses high quality feedback to give pupils constructive evaluation of their performance.	No or negative feedback.	Feedback about incorrect performance only.	Simplistic feedback about both correct and incorrect performance (good, wrong, yes, right, no, etc.).	Specific feedback and useful about both correct and incorrect performance.

Table 4. Pupils' Position on Abolishing the PLE from Their Essays in the 1995 PLE⁵³

<i>School Number</i>	<i>Abolish Exams</i>	<i>Not Abolish Exams</i>	<i>Unclear Essay or Other Position</i>	<i>Nothing Written on the Question</i>
1	10	7	2	3
2	17	8	1	0
3	15	21	0	0
4	4	14	17	5
6	4	9	2	5
7	0	1	6	12
Total for Luwero District = 163 Schools	50	60	28	25
10	40	20	1	1
11	7	18	2	0
13	4	8	15	3
14	2	7	27	10
15	31	33	0	0
16	3	9	9	3
Total for Masindi District = 253 Schools	87	95	54	17
Overall Total of Essays Reviewed = 416	137	155	82	42

⁵³Scripts were available from UNEB for 12 out of the 16 schools in the study.

**Table 5. Means and Standard Deviations for Classroom Observation Items
for 31 Classrooms in Uganda**
(8 Point Scale, from 1-8, where 1 is low and 8 is high)

<i>Classroom Observation Feature</i>	<i>Mean</i>	<i>Standard Deviation</i>
C1 Lesson Objectives Objectives of lesson are clear and well understood.	4.68	1.04
C2 Appropriateness of Activities Activities support lesson objectives.	3.07	0.81
C3 Elaboration and Examples Elaboration and good examples are used throughout lesson.	5.26	1.26
C4 Knowledge of Content Knowledge is accurate and thorough.	5.32	0.65
C5 Coherence and Content Flow Content elements of lesson fit together without disjuncture. Lesson contents form a coherent package.	5.13	1.26
C6 Process and Events Flow Continuity of process. Events of lesson form smooth sequence.	2.79	1.01
TP1 Opportunities for Practice Multiple opportunities for practice by each pupil.	3.64	0.98
TP2 Opportunities for Thoughtfulness Multiple opportunities for thoughtfulness by each pupil.	4.39	0.88
TP3 Cognitive Level of Learning Lesson emphasizes critical analysis.	3.94	1.03
TP4 Teacher Preparedness and Organization Teacher is well prepared. Functional elements are arranged for effective learning (teacher, pupils, work stations, materials, aids).	5.19	0.91
TP5 Pupil Participation Pupils personally participate in one of more activities (group work, manipulation of materials, etc.).	4.58	1.61
TP6 Questions Involving All Pupils Knowledge of all pupils is explored through teacher's questioning of individuals or groups.	5.29	1.07
M1 Quantity of Teaching Strategies A number of different teaching strategies are used to provide opportunities for pupils to become engaged in learning (lecture, dialogue, group work, projects. etc.).	3.79	0.70

<i>Classroom Observation Feature</i>	<i>Mean</i>	<i>Standard Deviation</i>
M2 Quality of Teaching Strategies The quality of teaching strategies enables pupils to become engaged in learning (dynamic presentations, questions, inquiry, rehearsing, examples, reinforcing activities, etc.).	4.59	1.14
M3 Pupil Attentiveness High percentage of pupils are on task, absorbed, interested, not daydreaming, or playing.	6.03	1.28
M4 Teacher Enthusiasm Teacher communicates interest, enthusiasm, and excitement about subject through voice, pace, nonverbals, etc.	5.74	0.93
M5 Pupil Enjoyment High level of enjoyment is shown by pupils.	4.64	1.11
M6 Pupil Initiative Pupils initiate learning activities and demonstrate creativity.	1.16	0.58
M7 Teacher Responsiveness to Pupil Initiative Teachers respond positively to pupil-initiated learning activities.	0.45	1.75
M8 Complexity and Interrelatedness Information is complex. Different points and perspectives are interrelated.	3.80	0.87
M9 Cognitive Challenge Level of learning positively motivates all or most pupils at appropriate level. Lessons are neither too easy nor too difficult for pupils.	5.83	0.82
F1 Quantity of Informative Feedback Teacher uses frequent feedback to give pupils constructive evaluation of their performance.	4.11	1.32
F2 Quality of Informative Feedback Teacher uses high quality feedback to give pupils constructive evaluation of their performance.	4.93	1.63

**Table 6. PLE Characteristics Linked to Teachers and Pupils in
Teachers' Interview Statements**
(Each entry is the number of statements made pertaining to that characteristic)

<i>Testing, Teaching, or Learning Characteristic</i>	<i>Characteristic Seen in PLE</i>	<i>Characteristic Seen in Teaching and Classroom</i>	<i>Characteristic Seen in Children's Performance</i>
Teaching Methods and Learning Strategies			
Questioning techniques and setting of questions changed (in general)	31	10	3
Change to new methods (general)	2	12	1
Practical, external use for learning. Linked to home. Applied. Relevant.	5	7	2
Multiple views and opinions explored.	3	3	0
Less emphasis on facts alone	3	4	0
Teach to exams at expense of learning. Exam-oriented.	1	6	0
Reasoning, thinking, explaining skills.	17	5	4
Creativity and self-expression.	1	1	3
Multiple teaching and testing methods.	2	5	0
Drill and practice.	2	6	1
Discovery and inquiry methods.	2	4	1
Participation, sharing, discussion, groups.	1	8	1
Individualized teaching and child support.	0	3	0
Problem-solving projects	0	3	2
Fewer multiple choice questions	4	1	2
Less lecture emphasis	2	5	0
Less use of direct questions	4	3	1

<i>Testing, Teaching, or Learning Characteristic</i>	<i>Characteristic Seen in PLE</i>	<i>Characteristic Seen in Teaching and Classroom</i>	<i>Characteristic Seen in Children's Performance</i>
Emphasis on understanding	6	9	3
Use of definitions, clearer information.	0	3	0
Expansion and wider knowledge.	1	2	2
Step by step learning and testing.	3	1	1
Child-centered education.	1	0	0
Fewer short and simple answers.	4	1	1
Less emphasis on rote memory.	6	1	0
Test-Taking Issues and Strategies			
Practice with testing and test-taking strategies.	0	7	1
Exams do not match the syllabus.	6	6	2
Cheating, exam leakages, malpractices.	2	0	2
Exams have trick questions.	2	0	1
Exams are more difficult than before.	4	0	4
UNEB more responsive to school issues.	2	0	0
Teaching Materials			
Current, updated materials	0	4	0
Change in textbooks.	0	1	0
Hands-on materials, aids, demonstrations.	0	7	0
Language Issues			
English improvement in children.	1	2	3
Language difficult.	1	2	1
Different style language used to teach and test.	4	2	0

<i>Testing, Teaching, or Learning Characteristic</i>	<i>Characteristic Seen in PLE</i>	<i>Characteristic Seen in Teaching and Classroom</i>	<i>Characteristic Seen in Children's Performance</i>
Less emphasis on grammar.	1	0	0
Fuller sentences, fuller answers, essays.	6	2	6
Fewer one word answers	1	0	1
Change in Teachers			
Teachers more resourceful	0	7	0
Teachers collaborate	1	2	0
Teachers better trained	0	4	1
Teachers preparing lessons, objectives	0	2	0
Teachers work harder	1	2	0
Teachers change	0	5	0
Teachers motivate	0	2	0
Change in Children			
Children motivated, attentive	4	1	4
Fear worry, stress	0	1	1
Increase in number passing. Children's performance improved	0	0	5
Urban schools do better than rural schools	0	0	3

Table 7. Summary of the Behavioral, Institutional, and Cultural Dimensions for the Uganda National Examinations Board Examinations Reform

<i>Dimension</i>	<i>Evaluative Status for the Reform</i>
Bounded Rationality	Teachers and headteachers know about the reform; teachers may not understand how to implement the reform in the classroom; no evidence of application in the classroom; no mechanism, except through teachers, to inform pupils; pupils and parents know far less about the reform; reform generated within UNEB and not fully or actively supported by the Ministry. The 'power of the exam' has spread the word; otherwise, minimal diffusion strategy—three newsletters, some workshops, and word of mouth.
Opportunism	Many problems noted; considerable concerns about cheating and advanced notice and preparation; urban areas have more information about changes, or so perceived. Many things done to reduce opportunism in the national examinations. Local exams or assessments would more vulnerable.
Asset Specificity	Better teachers in urban areas and upper grade levels; any assets from home are an advantage to a pupil; rural schools feel particularly stressed due to the lack of experience in the lives of pupils or teachers of this kind of examination. Those teachers and pupils better placed or capable will retain that advantage in the examinations.
Uncertainty	Teachers and headteachers learn about specific reform measures after the PLE examinations are administered. They can then use that information for the next preparation, but things may change again. Pupils are even less sure what to expect.
Number	One administration of the PLE per year; few selected for next level = high stakes.
Strange Loops and Complexity of Context	Single-dimension reforms exhibit strange loopiness; disregard of the complexity of the problems faced by education curtails progress or retards the management of the system. Many reforms increase loose coupling. Difficult to tell what the long-term implications of the UNEB reform are. It may have no impact on classroom instruction because teachers lack the skills to implement.

<i>Dimension</i>	<i>Evaluative Status for the Reform</i>
Institutional Symbols and Credibility	Exams mark the credibility of the system; changes risk differences in outcomes and ambiguities in attitudes about the quality of the system. Exams are clearly the defining feature of the schooling system; teachers try to follow exam perspectives. UNEB reform successful in changing the direction of instructional ideas but not the action. No challenges to the reform, either in terms of traditional values and symbols or credibility concerns.