

# **DECISIONS IN THE ESTABLISHMENT AND OPERATION OF A NATIONAL TESTING CENTER**

**Noel McGinn<sup>1</sup> and Massoud Moussavi<sup>2</sup>**

December 2010

*Acknowledgements:* This report was prepared with a grant from Open Society Institute. The authors would like to acknowledge the support of Hugh McLean and Aleesha Taylor, OSI London, in making this project possible. We would like to thank Michael Fine for his many ideas and suggestions and his careful reading of the section on corruption. We greatly benefited from discussions with Nazarkhudo Dastambuev, OSI Dushanbe, about the education system in Tajikistan. We are thankful to Carol Rosen for her careful editing and proofreading.

---

<sup>1</sup> mcginn@causalinks.com

<sup>2</sup> moussavi@causalinks.com

## Contents

**Executive summary 5**

*Suggestions for Open Society Institute 8*

**Introduction 9**

**A framework for decision-making 9**

**About this paper 11**

**A. Corruption in the process of selection for higher education 12**

**The nature and causes of corruption in education 12**

**The impact of corruption in selection and testing 13**

**Strategies to reduce corruption 14**

*Diagnosing the problems 15*

*Deciding on interventions 16*

**B. Methods used to select students for higher education 17**

**The variety of selection methods 17**

*Admission on graduation from high school 17*

*Standardized examinations 17*

**The effectiveness of selection methods 19**

*Comparative advantages and disadvantages among methods 19*

*Summary 21*

**C. Testing and assessment to improve the quality of education 23**

**What kind of information is needed? 23**

**The impact of assessments 24**

*Types of assessments 24*

<i>The pros and cons of centralized examinations</i>	25
<b>Examination results and school accountability</b>	26
<b>Testing and the quality of education</b>	26
<i>Secondary education</i>	27
<i>Higher education</i>	28
<b>D. Concluding remarks</b>	30
<b>The role of testing and assessment</b>	30
<b>The role of national testing centers</b>	31
<b>References</b>	32

## List of figures, tables, and boxes

### *Figures*

- 1. Diagnostic framework for the effects of input and process variables on the quality of learning 10**
- A.1. A simple model of teachers' behavior 15**
- C.1. A partial conceptualization of input and process variables affecting the quality of learning 23**

### *Table*

- A.1. Forms of examination fraud 14**

### *Boxes*

- A.1. Civil service examinations in China 12**
- B.1. University admission examinations in different countries 18**
- B.2. How different tests correlate with each other 20**
- B.3. Equity effects of admissions based on secondary school grades 21**
- C.1. The Texas assessment of academic skills 25**
- C.2. Early efforts to assess the quality of schools in the United States 28**
- C.3. Approaches to assessing university quality 29**

## Executive summary

Higher education is widely recognized as an important key to economic advancement. In many developing countries, however, university entrance examinations have been tainted by bribery and fraud. Resulting distortions reduce the practical social benefits from educational investment, as well as general public confidence in the system. A major role of the National Testing Center (NTC) in Tajikistan will be to establish and conduct a fair and transparent selection mechanism. In addition to preventing overt forms of corruption, a new NTC will have to deal with complex questions such as:

- How much control should government exert over the university admission process and how much should be left to the individual higher-level institutions?
- What kinds of tests are most appropriate to select qualified students?
- How can tests promote critical thinking and reasoning as well as factual knowledge?
- How can the objectivity, reliability, validity, and equity of the tests be ensured?

This study is designed to provide the NTC with baseline information about (a) the nature and scope of corruption in the education sector; (b) the potential impact of national testing on secondary school education; and (c) measures for achieving better equity in access to quality schooling.

A large literature has developed on corruption generally and in the education sector, addressing its many and varied forms, its social impact, and its antidotes. While there is basic agreement that corruption is harmful and should be discouraged, opinions vary widely on some key parameters: how exactly to define corruption, how to measure it, and how to mitigate its effects. Some take a narrow approach focused on overt conduct (such as paying a bribe to secure an economic benefit), with reform options that emphasize rules, enforcement, and incentives. Others view the problem in broader structural terms, with proposed solutions keyed to greater fairness, equality of access, and transparency. For purposes of analysis and planning, the important point for NTC is to understand these different perspectives and their implications for responsive action.

Corruption in education may not involve the lucrative kickbacks sometimes reported in other sectors, such as construction or petroleum, but its impact is significant because of its potential demoralizing effects on students and the general public. In their daily lives, most

people care about and are involved with schools more than they are with, say the petroleum sector. As a result, what happens in schools has a major impact on their sense of trust in the system and hope for the future.

The first part of this study provides an overview of corruption in the education sector, concluding that properly structured university entrance examinations can significantly reduce distortions in the selection process. We next survey practices that can be used to properly structure entrance examinations. This portion of the study has two parts, beginning with a review of common approaches for identifying the best students followed by a discussion of the predictive validity and comparative advantages of the different methods of selection.

Broadly speaking, there are two basic types of tests for entrance examinations: (a) “achievement” or curriculum-based tests that measure students’ knowledge of a particular curriculum and (b) “aptitude” tests designed to measure the cognitive skills and ability to learn. A further common distinction is based on the timing and frequency of testing, with one-time examinations for university entrance or a high-school diploma referred to as “high-stakes” tests because they affect important life outcomes and the routine measurements by teachers over the course of a student’s secondary education being “low-stakes” tests.

Research on testing methods has generally shown achievement tests to be a better predictor of overall student performance at university than aptitude tests, and this is reflected in the testing practices of most high-quality education systems. The most common approach in these systems has been to combine low-stakes internal examination results (or course grades) with high-stakes achievement testing. But this approach works well only in education systems that are free of corruption, in which teachers and administrators are well trained in assessment procedures and everyone generally supports the assessment process. However, when the results of low-stake tests cannot be relied upon because of rampant cheating and bribery or because the teachers are not well trained in assessment procedures, a well-designed centralized university examination administered securely and scored objectively and transparently will serve as a fair and inexpensive method for allocating limited space for new students.

Further points for NTC to consider are the effects of the centralized university examination on secondary education and the behavior of schools, teachers, students, and their families. These tests can increase accountability because they provide readily comparable measures of performance. Furthermore, if tests are designed to assess analytical and higher-order thinking skills, they can strengthen incentives for teachers to teach better and students to develop a deeper understanding of the subjects. Conversely, when tests are not comprehensive or encourage memorization, they can narrow teaching and learning approaches, leading to what has colloquially been called “teaching to the test.” In some

cases, standardization also has spawned a new service industry focused on preparing students for the tests.

The use of centralized examinations also bears directly on the equity of university admissions. A pitfall of admissions based solely on national tests (particularly aptitude tests) is that content may be inherently biased against poor or minority groups (for example, where testing measures cultural or other knowledge or experience not widely shared). Another common distortion occurs because poor students typically have fewer resources to prepare for such tests. Selection processes can be structured to address these concerns—for example, through measures that increase university attendance of disadvantaged groups without lowering the overall quality of the student body or subsidize test preparation services. Examples of good practice can be found in high-quality education systems for appropriate adaptation to Tajikistan's unique needs and circumstances.

The challenge for NTC will be to develop a university selection method appropriate for Tajikistan by tailoring the many examples of good practice in other countries to Tajikistan's particular goals and challenges. Choices among selection options will need to rest on an understanding of the relationships among many variables, such as student and teacher demographics and socioeconomic status, school organization, and parent involvement. It will also be important for NTC to assess the educational system's vulnerability to corruption and associated impacts on quality and equitable access. This paper proposes a diagnostic framework to analyze these critical relationships.

In undertaking these comprehensive assessments of Tajikistan's educational system and school population, NTC can also advance other educational objectives. These assessments would lay a factual foundation for interventions such as revising standards, creating incentives for better performance, and developing teacher training programs. By sharing this assessment information, NTC can have a broader impact on educational reform in Tajikistan beyond its decisions concerning the university selection methods.

National testing can be a valuable tool for ensuring the admission of high-quality students to universities, improving the quality of secondary school teaching, and reducing the distortions from bribery or other forms of corruption. It is only one part of a broader program to strengthen national education systems, however. The Open Society Institute may also wish to explore initiatives on educational accountability. Policy-makers should be accountable for their decisions on educational policies, rules, human resources, and budget issues; school principals and teachers for their decisions on curriculum and teaching techniques; students for their performance and behavior; and the general public for their engagement in the lives of their youth. Because most analyses so far have focused on accountability through external drivers (such as laws and their enforcement through penalties), an opportunity remains to foster accountability through public dialogues,

teacher training, student honor codes, and similar interventions that develop a strong social ethic and sense of personal responsibility.

### *Suggestions for Open Society Institute*

OSI might wish to consider, for example, an experimental program to evaluate and strengthen educational content and methods for students and the general public on issues of personal ethics, responsibility, and the value of hard work.<sup>3</sup> While such content has been developed in some other contexts, the quality has often been uneven or not readily transferable to other social settings. Furthermore, such an initiative might take advantage of particular cultural strengths and circumstances in a country such as Tajikistan—for example, by building on people’s strong religious beliefs, sense of community, and family honor and by relating those attributes to studies, work, and professional activities.

The Socratic approach and dialogue (that is, the OSI Debate Program) can also be extended to most subjects in the curriculum. For example, in mathematics, a way of developing analytical and higher-order skills is to learn how one can arrive at a solution to a problem using different methods; in history one might analyze the variety of causes for a particular historical event. OSI might want to consider creating incentives that would encourage students and teachers to pursue the Socratic approach to learning.

---

<sup>3</sup> A model can be found in the Debate Program that OSI initiated in 1995. DEBATE encourages critical thinking and personal expression.

## Introduction

The process of selecting students for entry to university can provide opportunities for cheating, bribery, and other forms of misconduct to secure an undue advantage. Such behavior, if widespread, can reduce the knowledge and skill level of the cohort that enters higher education and over the long term can lower the quality of graduates who take positions in society and the economy. It can also demoralize students, teachers, and the general public, reducing confidence in the selection process and civil society broadly. Corrupt acts by some can undermine an entire system.

A national testing center is an institution with authority to develop and administer the process of selection to higher levels of education. That vital role gives an NTC the opportunity to reduce the most overt forms of corruption like buying and selling of examinations, grades, and admission. Through more systematic assessment at selected grade levels and rigorous data collection, an NTC can also help to improve the quality of secondary and higher education and to reduce inequity in a country's education system. The primary focus of this review will be on an NTC's assessments and testing for entry to university and their impacts on reducing corruption in the admission process, improving the quality of learning and teaching, and increasing educational opportunity for disadvantaged students.

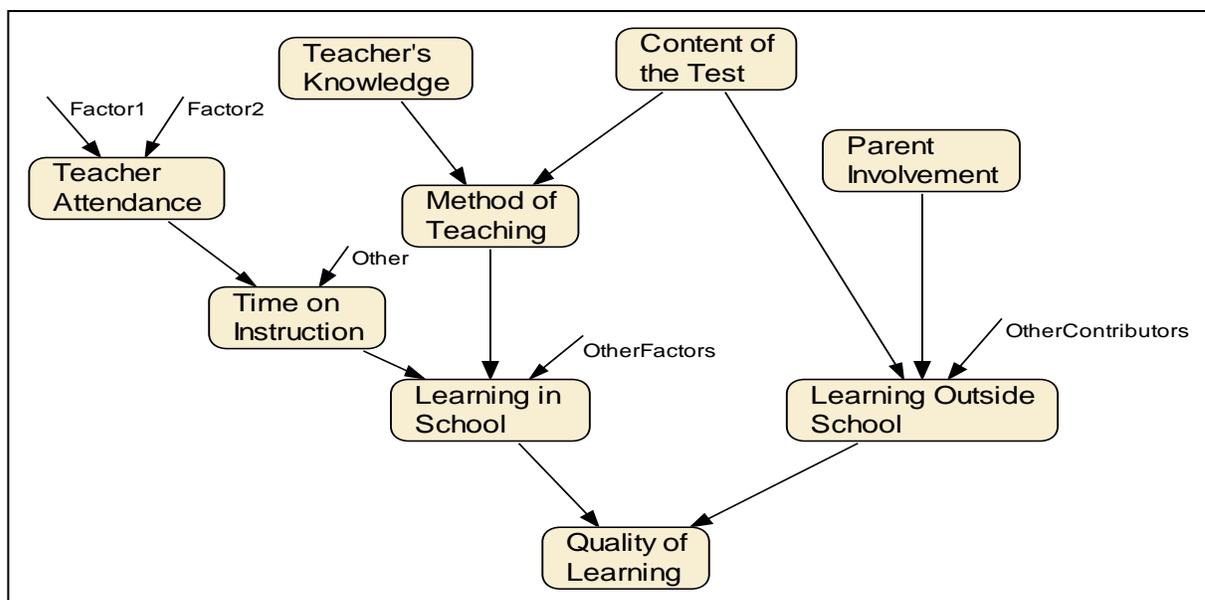
Decision-makers at an NTC face many questions: What selection methods are most effective? What relates most closely to the needs of stakeholders? What are the relative costs? What prerequisites are needed? What actions have the broadest impact? How have others addressed similar problems? They also need to consider how to address the issues raised as a result of the assessments: What processes and standards to put in place? What partnerships to build? How to ensure accountability? How to increase incentives for students to learn, teachers to promote critical thinking, and parents to become involved in the process?

### **A framework for decision-making**

Choosing the most cost-effective testing and assessment options involve tradeoffs. This paper offers a diagnostic framework to better understand the array of relevant factors, their interrelationships, and their impact on outcomes. For instance, with regard to corruption, these factors include the accountability of various actors (policymakers, teachers, students, parents, etc.), the existence of anticorruption laws and their enforcement, the level of information dissemination, the security of examination centers, and the incentives for misconduct.

Similarly, factors that affect quality of learning and teaching include the accountability of teachers and schools for results, the content and design of tests, the methods of teaching, teachers' training and academic background, students' social and economic status, parents' involvement, the adequacy of school facilities, and the attendance rates of teachers and students.

The diagnostic framework that we adopt develops such relationships and shows them graphically, where nodes represent the various factors and links represent the dependency relationships between factors (Figure 1).<sup>4</sup> This figure shows only a small set of input and process variables that affect quality of learning: how and what a teacher teaches are affected by teacher's training and knowledge as well as by content and design of tests; and method of teaching and time on instruction (among others) influence quality of learning in school. As seen in this figure, the effect of a factor (for example, the content of the test) on the quality of learning is mediated by a number of intervening variables. Understanding what a variable directly impacts is significant in correctly diagnosing a problem and selecting the right intervention.



**Figure 1. Diagnostic framework for the effects of input and process variables on the quality of learning**

We apply the same framework to develop an understanding of various factors that affect equity, or create opportunities for corruption, or can control it. For example, teacher attendance (or absenteeism, a prevalent occurrence in many developing countries) can be influenced by many factors including low wages, lack of recognition, and lack of

<sup>4</sup> Such graphs are commonly referred to as knowledge maps, causal models, or Bayesian networks. A Bayesian network has the additional feature in that the relationships are probabilistically quantified.

accountability. As in the graphical model for quality of learning, many factors are involved in understanding corruption. For instance, the quality of law enforcement affects accountability, while it is influenced itself by the severity of penalties and the reliability of enforcement.

Once this diagnostic framework is established, decision-makers can focus on interventions and evaluate their impact and cost-effectiveness. For example, a teacher training program is an intervention that can improve teacher's knowledge of teaching methods which instill higher-order thinking and reasoning skills in students. But as shown in Figure 1, tests also affect the teaching method that teachers use. When standardized tests encourage rote learning, teachers often revert to similar techniques rather than emphasizing analytical skills (World Bank, 2003). As a result, teacher training programs become less effective.

We note the importance of accountability as an underlying factor affecting all the outcomes. To have an equitable and high-quality education system, policymakers, staff in the implementing agency, teachers, students, and parents should all be held accountable for their decisions. But accountability is complex and multifaceted. For example, it would be wrong to hold teachers alone accountable for test results, as many factors contribute to a student's performance on a test (most important, family socioeconomic status), but teachers should be held accountable for their decisions about coverage of material, method of teaching, missing classes, etc.

### **About this paper**

This paper seeks to

- Identify the complex set of factors associated with corruption in the process of selection for higher education
- Suggest interventions to reduce corruption in higher education
- Describe the methods commonly used to select students to enter higher education
- Assess the reliability and validity of the different selection methods
- Analyze the potential of each method to improve the quality of education in secondary schools and universities and to enhance the equity of access to higher education for disadvantaged students.

We conclude with suggestions of possible strategies for implementation and research.

## **A. Corruption in the process of selection for higher education**

Corruption in the selection processes of public and private institutions is not a new problem: witness the efforts over 1300 years to eliminate it in civil service examinations in China (see Box A.1). Nor is the problem isolated. Corruption has been endemic across societies and over time. This scope and persistence should not discourage efforts at control, but rather caution against simplistic analyses and solutions. Indeed, efforts to combat corruption have risen dramatically in recent years, reflected in multilateral conventions, new national laws, and a growing volume of research and writing on corruption by academics, development organizations, and others.

### **Box A.1. Civil service examinations in China**

The Keju system for examination of candidates for civil service positions began in China in the year 606 and continued until 1905. This system was used to select persons for all positions in the Imperial government, from education commissioners to prime ministers. High scores brought financial rewards, social prestige, power, and fame. The examination process began at the local or prefectural level. Those who passed were allowed to take the examination at the provincial level and, if successful, could take the final examination in the national capital. Candidates were tested on their knowledge of Confucian classic texts, poetry, writing, and national policy issues. Both written essays and oral statements had to conform to detailed rules covering length, style, number of words, and format. Poems were judged on their rhymes, symmetry, tonal balance, and style. Although success at each level increased a candidate's prestige, only those passing the final examination were eligible for official appointments. By the end of the nineteenth century, only four in every million candidates might achieve this honor. Cheating and other forms of corruption were common during the whole period, despite many efforts to eliminate them (Suen & Yu, 2006).

### **The nature and causes of corruption in education**

How we define corruption determines our view of the nature and scope of the problem and the strategies to deal with it.<sup>5</sup> Some define it as overt conduct (such as paying a bribe to secure an economic benefit), with reform options that emphasize rules, enforcement, and incentives. Others define it in broader structural terms, in the context of fairness and equality of access. For example, if an admission exam measures math skills through urban examples (thus, unfamiliar to a student in a rural area), is its bias because of corruption or poor design? In fact, undesirable outcomes may often be due as much to lack of capacity

---

<sup>5</sup> Appendix A.1 provides a detailed survey of various definitions of corruption, its forms and faces in education, and general strategies for combating it.

and poor management as to overt corruption. However, a test badly designed or poorly supervised often creates opportunities for corruption. In this context, corruption can be viewed as an outgrowth of poor governance. (Lewis & Pettersson, 2009).

As international attention has focused more on governance issues in the past 15 years, a dramatic change in attitudes toward corruption has raised scrutiny in all sectors, including education. Recent changes in international and national laws have promoted integrity and transparency; funders and civil society organizations, such as the World Bank, the Open Society Institute, and Transparency International), have embraced a broad anti-corruption reform agenda.

### **The impact of corruption in selection and testing**

Reports on corruption in education confirm popular perceptions of its wide-ranging ill-effects on students, teachers, the education system, and society as a whole. The immediate impact of a corrupt act is obvious—one individual or group suffers, another benefits. Less visible, however, are long-term consequences on employment, income levels, and international competitiveness. Although a number of econometric models demonstrate an association between corruption and economic outcomes (Weidman & Enkhjargal, 2008), supporting hard data are currently limited. Similarly, to our knowledge no formal studies have yet been conducted showing the causal relationship between corruption (in the form of denying access to a qualified applicant or failing a student who has met academic requirements) and individual life outcomes.

Another critical impact of corruption in selection and testing is the demeaning and demoralization of a country's educational system. The case of fraud in examinations illustrates the potential for harm to the integrity of academic evaluation. Fraud can occur at various stages in the examination process (Table A.1). Examples include the sale to students of the questions to be asked in oral examinations (Italy), the hiring of a substitute to sit an examination (China), the bribing an official to alter the score obtained on an admission test (India), and the use of "invisible" microphones and earphones to communicate with someone outside the examination center (United States).

Such cheating has many motives. The expansion of education has deflated the value of a given amount of schooling: the next generation requires higher levels of education to enjoy the same social and economic status as their parents. More students apply for admission to university, in some countries greatly exceeding the number of available places. With much riding on the result of an entrance examination, the likelihood of corrupt behavior increases. At the same time, persons controlling examination contents, administration and scoring are tempted to supplement low incomes by demanding bribes. Corrupt actions are more likely in conditions of poor administration and supervision.

**Table A.1. Forms of examination fraud**

<u>Stage in process</u>	<u>Form of malpractice</u>	<u>Actors</u>
Examination development	Leakage of questions	Test and subject specialists
Printing and distribution	Selling questions	Clerks, printers
Preparation of students	Tutoring, cramming	Teachers
Administration of examination	Impersonation	Students
	External assistance	Teachers
	Use of banned materials	Supervisors
	Copying	
	Collusion	
	Substitution of scripts	
	Improper assignment	
Marking of examination	Various malpractices	Teachers
		Clerks Graders
Awarding Diplomas	Manipulation of results	Administrators
	Fake degrees and diplomas	Teachers Clerks
Use of results	Paying bribe to enter University (regardless of Exam results)	Students
		Administrators Clerks

Source: Adapted from Hallak & Poisson, 2007, p. 236.

### **Strategies to reduce corruption**

Building an anticorruption system is like designing a healthcare system.<sup>6</sup> In healthcare, before deciding on how to allocate resources to various options for prevention and treatment (such as educating the public, building more hospitals, or training more doctors and nurses), one would need to first clarify the objectives (such as reducing infectious diseases or infant and maternal mortality), and understand the causes of problems. We suggest taking a similar approach for dealing with corruption. Thus, we treat anticorruption strategies as a decision problem. Like any decision problem, there is a diagnostic phase,

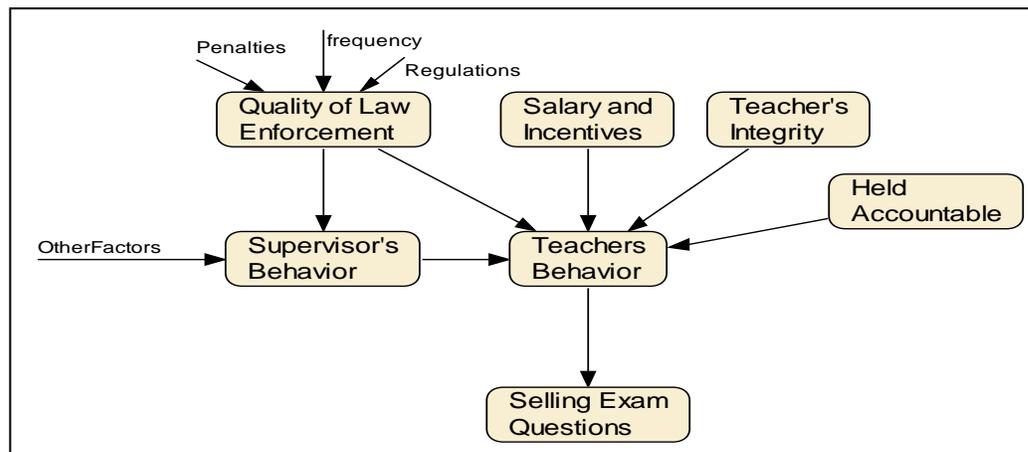
---

<sup>6</sup> Klitgaard (Klitgaard, 2000) has used the analogy of disease pandemic for corruption. Not an ideal metaphor in our opinion.

which involves identifying the vulnerabilities, causes, and sources of the problem, and a decision phase, which involves selecting the most cost-effective option.

### *Diagnosing the problems*

In developing a strategy to reduce fraud in examinations, for example, inherent vulnerabilities and risks have to be assessed at each stage of the process shown in Table A.1. Consider the *administration of examination* stage. The vulnerabilities might include clarity of rules and guidelines; level of information dissemination; quality of law enforcement; severity of penalties; properties of the examination centers and their security level; identification control; various attributes of students, teachers, and supervisors, such as their capacity, training, background, level of education, and wages; and parents' involvement in the process. Using the diagnostic framework discussed in the Introduction section, these factors and how they influence each other can be visually represented in a schema of nodes and links (Figure A.1).



**Figure A.1. A simple model of teachers' behavior**

Figure A.1 depicts relationships among several factors, based on knowledge and experience in the field. It shows some of the important influences on a teacher's behavior. They include the teacher's integrity, his or her salary and incentives, the supervisor's behavior, the level of accountability, and the quality of law enforcement. Other nodes in the graph are in turn influenced by some other factors.<sup>7</sup> For instance, the quality of law enforcement is affected by the frequency of enforcement, the likely penalties, and the existence of good regulations. The graph also presents selling exam questions as one of the possible outcomes of teacher's unethical behavior.

In most situations, there are many more factors to consider, and an in-depth diagnosis may need to delve into many country-specific conditions (the state of regulations, efficacy of

<sup>7</sup> Of course not all the factors or relationships are shown in this simple example.

courts, recruitment and training of employees, etc.). Such a graphical representation is a good way to understand the underlying relationships as a step toward diagnosing specific problems. Accurately identifying those problems (low wages, lack of training, poor law enforcement, etc.) is the critical prerequisite to determining the right interventions.

### *Deciding on interventions*

Once a range of control mechanisms, reforms, or other interventions are identified, the alternatives can be ranked in order of their cost-effectiveness. Some countries, especially in Asia, emphasize moral education, there are no studies designed to demonstrate that training in ethics can significantly reduce corruption.<sup>8</sup> In the United States there is a long tradition of research on the effectiveness of university honor codes.<sup>9</sup> According to one study honor codes have a significant influence on students' behavior, if they are supported by faculty and staff (McCabe, Treviño, & Butterfield, 2001). However, a review of studies over time suggests that cheating is prevalent and on the increase in U.S. institutions.<sup>10</sup>

The establishment of a national testing center can also help to reduce exam fraud. An NTC can facilitate the control of test construction, distribution, and scoring, and it can clearly define rules and procedures. Computerization of scoring can ensure equal treatment and rapid distribution of results. Well-paid and highly trained NTC staff can be isolated from pressures. The centralization and transparency of the process makes it easier to identify corrupt acts and actors.

The success of Georgia's national testing center in fighting corruption in higher education is instructive. In 2005 Georgia passed a new law that made unified examinations mandatory for all applicants, and implemented it effectively by establishing control mechanisms and increasing transparency. Control mechanisms included ensuring confidentiality and fairness of grading by identifying candidates through barcodes on their exam papers, printing examination booklets in a secure overseas facility, and equipping the testing centers with surveillance cameras and video monitors that let students' relatives observe the examination process (World Bank 2003). These efforts gained public trust. The success of the new system in Georgia owes much to the strong political will of the new government, the capacity of the staff of the center, and the sufficient financing of the center (Gabrseck, 2010).

---

<sup>8</sup> Although no dedicated studies have been done, most observers (and also community members) believe intuitively that an ethical grounding can positively influence attitudes and practice. This might be a fruitful area for further OSI inquiry.

<sup>9</sup> Most of the research evaluates the effectiveness of the code at a particular institution. For a bibliography see <http://wrt-howard.syr.edu/Bibs/HonorCodes.htm>.

<sup>10</sup> This is illustrative of the benefit of analysis of multiple contributing factors. It may be that there are positive impacts from an honor code that in tougher economic times or environments can be swamped by countervailing factors.

## **B. Methods used to select students for higher education**

Countries and universities now use a variety of methods to choose among secondary students applying for higher education. These methods profoundly affect the students' life prospects, but they may also broadly affect the country's quality of secondary and higher education, as well as its climate of corruption or inequity. Unfortunately, there have been no controlled experiments to determine which methods of selection are more effective in reducing corruption or improving learning. Instead, researchers have focused on the immediate effectiveness of selection methods in identifying qualified candidates or predicting students' success in university courses, and their impact on teaching practices and incentives to cheat.

### **The variety of selection methods**

High demand for higher education, requiring selection from among large numbers of applicants, is a recent phenomenon. When few students sought entry, it was possible to use personal methods of selection. Today, to contain costs and ensure integrity, many countries have introduced multiple-choice examinations and computerized scoring. Each method of assessment has benefits, but also drawbacks. A review of university admission practices around the world suggests several approaches in play (Helms, 2009).

#### *Admission on graduation from high school*

In a few countries, graduation from an accredited high school is sufficient for admission (see Appendix B.2 for a definition of accreditation and a glossary of terms). Those countries of the European Union that are signatories of the 1997 Lisbon Convention will admit for university study any person who is qualified for university admission in their own (EU) country. Finland, Norway, and Sweden, for example, qualify their students for admission on the basis of upper secondary or high school grades; those qualified students can then enter university in any of 17 or more European Union countries. In the United States a few private universities make admission decisions principally based on high school grades, letters of recommendation, and personal statements.

#### *Standardized examinations*

Most countries, however, administer some kind of standardized examination to assess applicants' knowledge and skills. Such tests are not new. They were first employed in China in the sixth century, and the University of Paris (Sorbonne) had its own entrance examination in the thirteenth century (Office of Technology Assessment, 1992).

Today there is growing enthusiasm for the use of standardized testing not just to evaluate students or select applicants to university, but also to assess teachers, schools and national systems (Kellaghan, 2004). More than three-quarters of all developed countries and half of

all developing countries have conducted some form of national assessment of student knowledge and skills (Benavot & Tanner, 2007). Two forms of externally administered, standardized examinations are widely used for university admissions: those designed to measure the past learning of students (achievement tests) and those that intend to measure applicants' ability to learn in the future (here labeled ability tests).

**Achievement tests.** With the development of national systems of education in the eighteenth century secondary education emerged as a means of preparation for higher education. This led to the development of examinations to ensure that students reached required levels of knowledge and skills. Now students in many countries face achievement tests to enter university (see Box B.1).

#### **Box B.1. University admission examinations in different countries**

Examples of achievement tests around the world include the General Comprehensive Examination, or GCE, administered in the United Kingdom, Hong Kong, Malaysia, Pakistan, Singapore, and Sri Lanka; the Abitur examination given in Germany, Finland, and Estonia; and the Matura employed in Albania, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Hungary, Italy, Liechtenstein, the Republic of Macedonia, Montenegro, Poland, Serbia, Slovakia, Slovenia, and Switzerland. In South Africa admission to university is based on scores on the Matric examination (Centre for Higher Education Transformation, 2005). The United Kingdom has two GCE examination boards (or test centers), one for England, Wales and Northern Ireland, another for Scotland. In Germany each of the states has an examination board (see Helms, 2009; Webber, 1989). In France's centralized system, all candidates are tested with the same questions, but scoring has not necessarily been uniform. Most examiners have been locally recruited and received little training in how to evaluate student responses. Examination results have been used only to evaluate students, not schools or the education system (Zwick, 2007). In China and Iran university candidates take a centralized national examination that covers the subjects taught in high school.

**Aptitude tests.** A student's ability to learn is less easily measurable than his or her mastery of a particular body of knowledge or level of achievement, so aptitude tests measure the student's relative ability to learn, compared with other students. The use of such tests in university admissions appeared in the United States following the rapid expansion of secondary education in the second half of the nineteenth century (see Appendix B.3). Like achievement tests of the period, these early aptitude tests had the effect of stimulating high schools to improve their teaching and universities to make their requirements for admission more uniform (Barker, 1967).

Most universities around the world use information from several sources to select among candidates for admission. In addition to test scores, most use secondary school grade point averages; many require letters of recommendation; most require a writing sample, often in the form of a personal statement; and other sources such as work and life experience. In the

United States although 90 percent of the institutions require that applicants take either the SAT or the ACT test, only 70 percent of the institutions routinely take scores into account.

### **The effectiveness of selection methods**

Should admission tests be based primarily on subject matter knowledge or on anticipated learning ability? The founders of a national testing center must ultimately decide this question for themselves. Several factors have to be taken into account in making that decision:

- The superiority of one kind of test over another, as discussed in Appendix B.4.
- The capability of a given national testing center to construct either kind of test.
- The cost of designing, constructing, administering, and scoring each kind of test
- The impact of each kind of test on the quality of education in secondary schools and in universities.
- The potential contribution of using a particular kind of test to corruption and inequality in the education system.

Most countries in the world now rely principally on measures of secondary school knowledge (grade point averages, school diplomas, or subject-based examination scores) to make admission decisions.<sup>11</sup> The predictive validity of these methods is reviewed in Appendix B.5.

### *Comparative advantages and disadvantages among methods*

Almost all the research indicates that secondary school grades are a better predictor of university success than scores on subject-based examinations or scores on ability tests (see Appendix B.5). Furthermore, grades, which are much less closely correlated with students' socioeconomic characteristics than standardized test scores, are less subject to bias and thus have the potential to reduce inequities in the selection process. And because grades are compiled over a longer timeframe than a high-stakes testing environment, they may be less likely to spur cheating. Grades, too, have shortcomings, as discussed below, and may be most effective when a national testing center uses them in combination with testing methods to ensure that university selection processes are educationally sound, unbiased, and ethical.

---

<sup>11</sup> It may be significant that the list of countries includes those where students score highest on international tests, such as that of the Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA).

***The potential to predict academic success.*** In all the studies reviewed, secondary school grades correlated more highly with first year university grades than did scores on achievement tests given at the end of secondary school or scores on aptitude tests.<sup>12</sup> At the same time, the three sets of scores correlated moderately with each other (see Box B.2).

**Box B.2. How different tests correlate with each other**

SAT scores correlate ( $r = 0.50$ ) with scores on A-level tests (McDonald A. S., 2001). When combined in a multiple correlation or multiple regression equation, secondary school grades were always the most powerful predictor variable. For example, in a large sample of students taking the SAT in the United States, high school grades alone had an adjusted  $r$  value of 0.54, while SAT alone had an adjusted  $r$  value of 0.51. Adding SAT scores in the equation increased the  $r$  value from 0.54 to 0.62 (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). A study in Sweden obtained secondary grades and SAT scores for 7000 first-year engineering students and compared to their credit points (grades) in university. When entered in a multiple regression equation secondary grades yielded a regression coefficient of 0.47 compared to 0.07 for SAT and 0.04 for a nonverbal problem-solving test (Gustafsson, 2003).

Subject examinations appear to be the second best predictor, perhaps because their content is more closely related to the content of subjects taken in the first year of university than is the content of an aptitude or general knowledge test. For example, A-level examinations in the United Kingdom are better predictors of undergraduate medical school performance than are aptitude tests, most likely because of the linkage between A-level exam content and medical school courses.

***The potential to reduce inequity.*** At least in the United States, school grades are a less biased predictor of success than are ability tests (see Box B.3 for one study). The use of norm-referenced<sup>13</sup> selection tends to favor some ethnic groups over others when admission to university is highly selective. Criterion-referenced selection, in contrast, increases the proportion of qualified members of disadvantaged ethnic groups admitted to university. Socioeconomic status has a moderate to high correlation with academic achievement in elementary and secondary school and its effect is cumulative over time, so that the gap between average achievement of poorer students and wealthier students is larger in the upper grades. Ethnic groups from lower income groups are therefore less likely to score at the upper level in admission examinations. When the space is limited, very few of the low-income ethnic group will enter university. This tendency replicates the conditions that produced the original status of inequality.

In the United States, for example, African American students score significantly (one standard deviation) lower on the SAT than do whites (McDonald, Newton, Whetton, & Benefield, 2001). As a consequence only a small proportion of African-Americans enter

<sup>12</sup> See Beatty, Greenwood, & Linn (1999) for one perspective on this question.

<sup>13</sup> See Appendix B.2 for the glossary of terms.

and complete university. Nonetheless, African-Americans admitted to university obtain higher grades than predicted by their SAT scores, while white males obtain lower grades than predicted. Grades are less related to socioeconomic status or ethnicity of the student than are SAT scores (Geiser & Santelices, 2007).

### **Box B.3. Equity effects of admissions based on secondary school grades**

One study tracked the academic performance over four years of almost 80,000 students who entered the University of California between 1996 and 1999. They found that the superiority of the high-school record over standardized tests extended beyond its predictive value alone, insofar as prediction was possible, and HSGPA had other important advantages as an admissions criterion. High-school grades were much less closely correlated with student socioeconomic characteristics than standardized tests. Within the UC sample, HSGPA was only weakly correlated with family income, parents' education and school API [Academic Performance Index] rank, whereas SAT scores bear a strong, positive relationship to each of these measures. As a result, HSGPA had a less adverse impact than standardized admissions tests on underrepresented minority applicants, who come disproportionately from disadvantaged backgrounds." (Geiser & Santelices, 2007, pp. 25-26).

*The potential to reduce corruption.* A national testing center has the potential to reduce corruption and to improve the level of academic performance of students entering higher education. As discussed earlier (section A), corruption has already declined in several of the countries that recently established centers.

#### *Summary*

Although grades are the better predictor, the studies also show that the three methods reviewed--course grades in secondary school, achievement tests, and aptitude tests--are measuring something in common. Those who champion the use of secondary school grades as the primary factor for making admission decisions argue that past performance is the best predictor of future performance because grades are awarded by a number of different teachers over a number of years, while tests are one-shot events, and grades reflect motivation and persistence as well as learning (Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; Richardson & Abraham, 2009).

At the same time, high school grades are criticized as a measure for selection among university candidates because they are subject to corruption, not always reliable and not comparable across schools because assigned according to different criteria, and so on. Furthermore, in many countries there has been a steady inflation in grades during the past 20 or 30 years. In a highly unstable political climate it may be impossible to use high school grades as a selection criterion.<sup>14</sup> Most teachers have received little training in

---

<sup>14</sup> Some research has shown that rank in class is a better predictor of university performance than are standardized tests (Niu & Tienda, 2009). Class rank reflects the level of effort a student has made in his or

assessment, and schools and school systems provide few guidelines for assigning grades. The grading process is highly subjective and influenced by factors such as personal relationships unrelated to students' actual knowledge. Especially in countries with low levels of political stability, grades may be so unreliable that they cannot serve as valid predictors of success in university.

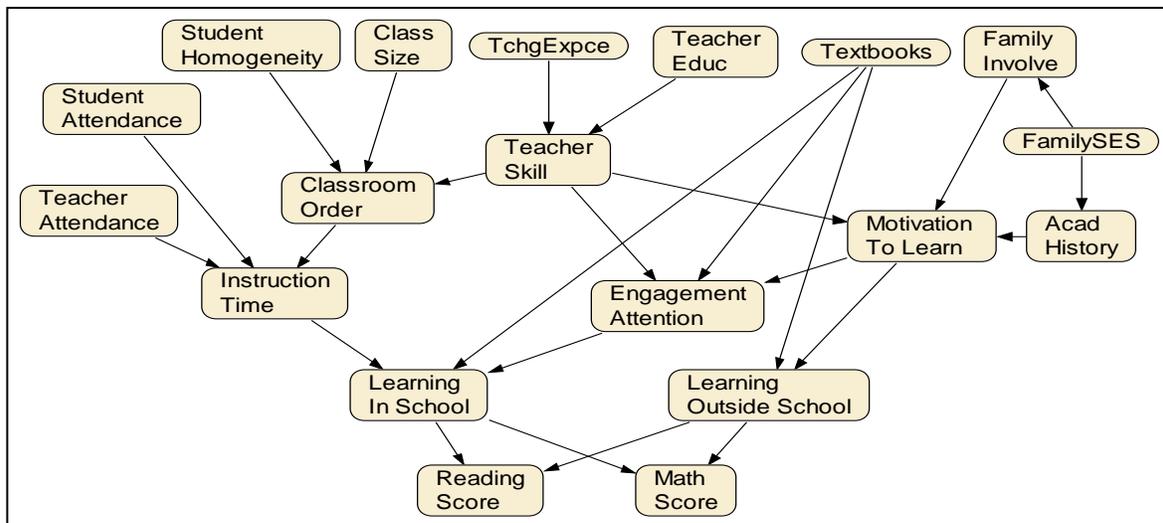
External subject-based examinations are a reasonable compromise solution. They can be controlled and monitored in the same way as ability or aptitude tests, while providing a much more comprehensive assessment of student knowledge. Their assessment of student knowledge has two benefits: university curriculum for the most part assume that students begin their higher studies knowing what was taught in secondary school; and the individual factors that contribute to learning at the secondary level (motivation, ability, health, etc.) are likely to continue in university. University departments can make selection decisions based on examinations linked with their subject matter (math and physics for engineering departments, for example).

## C. Testing and assessment to improve the quality of education

Improving the quality of education—the fundamental variable in the learning process—remains a primary goal of education reform. Yet it remains an elusive goal to quantify. Quality of education is usually defined in terms of inputs (textbooks, teachers’ education and experience, students’ academic background, family status, etc.) and process variables (instructional methods, time spent on instruction, students’ engagement, etc.) that contribute to learning. Efforts to benchmark current levels of quality or evaluate progress thus depend on gathering information for measurement.

### What kind of information is needed?

Cognitive psychologists are concerned about the learning process and factors such as student motivation, attentiveness, and intelligence; educators are concerned primarily with the teaching–learning process, with the actions and cognitive processes of both teacher and student; economists consider school as an organization for the transformation of resources. Each profession has a different perspective on the kind of information needed to assess quality of education (see Appendix C.1 for a detailed discussion). Each perspective in fact complements the others, and each is affected in complex ways by an array of input and process variables, as suggested in Figure C.1 ( Moussavi & McGinn, 2010).



**Figure C.1. A partial conceptualization of input and process variables affecting quality of learning**

## The impact of assessments

A national testing center can help to improve the quality of secondary education and higher education institutions by collecting data on a number of these variables through university entrance exams, sample-based assessments, and summative assessments at selected grade levels.

### *Types of assessments*

*Summative* (or *outcome*) assessments use standardized examinations to measure the level of performance of students or schools and are usually carried out at the end of a year or cycle. They are relatively inexpensive to conduct and can help determine whether some kind of intervention is required. But they are less helpful in determining which intervention would be most effective.<sup>15</sup>

To determine the most effective interventions, assessments should be carried out frequently to give teachers timely feedback on whether a particular student or class learned the material and whether a particular method or form of presentation was effective. This kind of assessment is called *formative* (or *implementation*). It provides information on the temporal connection between a given instructional practice and its effects.

To judge the success of their operations, education systems need information from both formative and summative evaluations. Standardized curriculum-based tests (including university admission tests) generate data that can be used for both summative purposes and, if the tests are given frequently, for formative evaluations. Ability and aptitude tests, as well as cumulative grade point averages used for admission decisions, produce data for summative evaluations.

Examination should not be the only basis for evaluating the effectiveness of secondary schools, either individually or comparatively. While average test scores reflect the knowledge level of students, they do not indicate where and when that knowledge was obtained. Evidence from locations as diverse as rural Kenya and urban Chicago shows that holding schools and teachers accountable solely on the basis of examination results is not a good idea. In that type of accountability regime, scores may rise, but teachers tend to teach to the test and manipulate the results (World Bank, 2003).

A better indicator for assessing the contribution of a given school to a student's knowledge is the difference between test results at the end and beginning of the year—so-called *gain scores*. Similarly, a better assessment of an individual teacher's performance is the gain score of test results in the subjects taught by the teacher.

---

<sup>15</sup> These assessments can also include information on costs associated with schooling, but few education systems in the developing world have disaggregated the cost of running a school, and even fewer know how that cost is associated with learning (World Bank, 2003).

### *The pros and cons of centralized examinations*

Centralized examinations have advantages and disadvantages. On the positive side, they do increase accountability. Because information about tests are readily available, parents get more involved and will put pressure on their children as well as schools for better results, and students prepare better. Centralized curriculum-based exams have a positive impact on student performance, as large an impact as parental education, according to one study (Woessmann, 2003). But improvement in performance in one kind of test may not necessarily mean deeper learning or greater knowledge if the test does not measure higher-order skills and thinking or if it has a limited coverage of the curriculum (see Box C.1).

On the negative side, an unavoidable consequence of high-stake tests, such as university admission or high-school exit tests, is that schools are often expected to produce ever-improving results and consequently teachers teach to the test. When the tests are not well designed, teaching to the test has negative consequences. Examples of poorly designed tests are the ones that emphasize only parts of the curriculum, or tests that promote rote memorization rather than higher-order skills. That makes light of the work of teacher training programs that attempt to instill the pedagogical techniques that promote critical reasoning and thinking (World Bank, 2003).

#### **Box C.1. The Texas assessment of academic skills**

Texas was one of the first states to use a standards-based examination, the Texas Assessment of Academic Skills (TAAS) to hold schools accountable for meeting standards. The test was introduced as part of a statewide reform of secondary education begun in the early 1990s, which set high goals for improvement of student learning. Studies showed that it had high reliability and correlated well with other tests of academic ability such as the Scholastic Aptitude Test (SAT) and the National Assessment of Educational Progress (NAEP). As the reform was implemented, scores on the TAAS climbed steadily. By the late 1990s the government of Texas was reporting very large average gains in test scores. The governor of Texas later claimed that the success of the educational reform was a major factor in his election to the presidency of the United States.

In reality, however, the improvements in TAAS scores did not reflect improvements in the general knowledge of students in Texas. Studies showed that while average TAAS scores had increased over time, no equivalent increase had taken place in either average NAEP or SAT scores. Nor was there a significant improvement in the proportion of students graduating from high school (Haney W., 2000). Further research revealed that school officials in Texas had elevated TAAS average scores by excluding some students (principally those in special education programs and Spanish speakers) from testing (Heilig & Darling-Hammond, 2008).

Teaching to the test also adversely affects minority students. In a study in the United States, teaching to the test was found to be most common in low-income school districts and least common in high-income ones (Amrein & Berliner, 2002).

Teaching to (and learning for) the test are likely to occur as long as there are high-stake tests, where a student's future may hang in the balance. What is important is to reduce negative consequences by designing tests that are reliable, fair, and equitable assessments of skills and knowledge intended by the curriculum. There is some evidence that the introduction of high-stake tests in some schools prompted students to cheat and teachers to give unauthorized assistance to students (Jacob & Levitt, 2003). However, these consequences must be balanced against evidence that high-stake testing is an effective mechanism to help reduce corruption in the selection process for university. In deciding on a centralized test, a national testing center has to be concerned with the tradeoffs between the cost of the test, its reliability, and its potential impacts on equity and corruption.

### **Examination results and school accountability**

In analyzing school accountability three important issues have to be considered (World Bank, 2003):

- *What else besides the school has influenced results?* The attributes of students, their peers, and their families are the most important contributors to student performance.
- *How reliable are the results?* An unavoidable feature of statistical sampling is that schools with a small number of students have large variability in results.
- *What are the consequences of good or bad results?* Education systems, schools, teachers, parents, and students all have a stake in the outcome and must decide whether to reward good performance, intervene following bad performance, or both.

What is certain is that school accountability is a complex matter, and schools cannot be held accountable solely based on the results of standard examinations. Appendix C.2 provides a detailed description of the issues surrounding high-stake tests, their impact on school accountability, teaching to the test, equity, and the potential for cheating.

### **Testing and the quality of education**

Educators who adopt large-scale testing as a means to improve the quality of education tend to see teaching and learning as a stochastic process requiring continuous assessment and correction. In that view, administrators and teachers have to be professionals capable

of autonomous decision-making, rather than skilled workers capable of following rules or recommendations prepared by others.

Formative assessments can provide information on many of the factors that are depicted in Figure C.1. For example, if data is collected for inputs such as classroom order, teacher attendance, student attendance, and instruction time, we can then develop better understanding for how each of the inputs affects the intermediate factor *instruction time*. Once a model of the type shown in Figure C.1 is developed, it can be used as useful tool for diagnosing problems and deciding on the most effective interventions.

### *Secondary education*

Standardized tests have been used to evaluate individual students for many years, but only recently have they been used to assess the quality of educational systems, to identify effective and ineffective teachers, and to identify more effective methods of instruction (Box C.2). Consequently, there is relatively little published research on how best to use national testing to improve secondary schools and even less on how to improve instruction in universities.

Early studies focused primarily on how teachers use data generated by the assessments they carry out using multiple-choice tests (Yeh, Herman, & Rudner, 1981). A more comprehensive work describes the use of assessment for school improvement in a number of countries (Kellaghan & Greaney, 2001). In the United States assessment data are bundled with other kinds of information (demographic, economic, student records) to support a process of data-driven decision-making by policymakers, administrators, and teachers. Test scores are used primarily to set improvement goals and targets and to monitor schools and teachers. In some instances schools use assessment data to identify problems in specific parts of the curriculum and misalignments between the content of school-based assessments and state or national assessments. Some teachers use the large-scale data to assess their own performance and to indicate where they should strengthen content knowledge or teaching skills (Marsh, Pane, & Hamilton, 2006).

Many factors influence teachers and administrators' use of data for decision-making:

- ***The accessibility of the data.*** This involves not just electronic access to online data banks but also training in the use of analysis procedures.
- ***The perceived (and real) quality of the data.*** Improving the public perception of the national testing center as a professional and incorruptible organization will also contribute to greater use of the data generated.

- ***The motivation to use the data.*** Teachers and administrators may require some urging to examine assessments of their performance. Incentives can contribute to motivation, but may also lead to corrupt actions.
- ***The timeliness of the data.*** As noted above, we learn most effectively when we receive immediate feedback about the results of our actions. Test scores at the end of 12th grade will contribute less to the improvement of quality than will test scores at the end of each grade.
- ***The capacity and support of staff.*** Teacher training programs typically include nothing on how to use data to improve teaching. Administrators will have to provide teachers with opportunities for professional development.
- ***The press of other obligations.*** Lack of time is frequently a limiting factor. Training and analysis are time-consuming. Improvement of the quality of education is a lengthy process (Marsh, Pane, & Hamilton, 2006).

#### **Box C.2. Early efforts to assess the quality of schools in the United States**

Until 1966, no country had carried out a quantitative assessment of its education system using rigorous methods of measurement of students' knowledge and skills. In that year James Coleman and colleagues published the results of a national sample survey of 150,000 students commissioned by the Congress of the United States (Coleman, 1966). The objective of the study was to assess the equality of educational opportunity across racial groups, comparing the quality of their schools as indicated by measures of physical inputs, characteristics of teachers, and student levels of knowledge and skills measured by standardized tests. The 700-page report claimed that differences in quality of schooling had little effect on student knowledge. Family and socioeconomic factors were thought to be more important for explaining why some students knew much more than others. This unexpected conclusion provoked an ever-expanding flood of studies attempting to define the factors that explain individual learning outcomes and school quality.

Some work has been done on the development of programs to promote the use of data to improve educational quality (Boudett, Murnane, City, & Moody, 2005; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). One comprehensive source of studies on data use to improve school quality is Gowalski & Lasley (2009).

#### ***Higher education***

In the United States, as competition has increased among universities for financial resources, pressure has also come to improve the quality of higher education. International concern for institutional quality has also risen, as a result of globalization of both labor markets and educational markets. External quality assurance is proposed as a means to assess the equivalence of university studies in different countries (Billing, 2004). This is a particular concern in the European Union, but also in Africa and Asia where markets are

globalizing. Most large universities now have an office of institutional research that is responsible for assessing the quality of teaching. Once procedures for measuring institutional quality have been developed and validated, researchers can compare the effectiveness of different approaches for improving the quality of higher education. A summary of various approaches to assess university quality both in the United States and internationally is provided in Box C.3.

**Box C.3. Approaches to assessing university quality**

Several methods have been developed for assessing university quality (Chun, 2002). One approach calculates a quality index based on graduation rates, racial and ethnic composition of the student population, endowment, student to faculty ratio, qualifications of faculty, admissions test scores, and other sources. This could be considered an "input" definition of quality. A second approach asks "experts" (higher education faculty and administrators) to rate the quality of institutions on dimensions such as quality of faculty and research. A third approach used in some countries relies on surveys of students. Finally, some effort has been made to assess the learning outcomes produced by different higher education institutions.

By 2005, 27 of the 50 United States had established some measure of learning outcomes to assess the quality of state universities (Klein, Kuh, Chun, Hamilton, & Shavelson, 2005). Based on the experiences of the states, researchers have developed a College Learning Assessment approach that directly measures student progress in critical thinking, analytic reasoning, and writing. CLA scores are correlated with students' entering SAT scores (Klein, Kuh, Chun, Hamilton, & Shavelson, 2005). The approach emphasizes that the CLA can be used as a formative assessment device for the continuous improvement of teaching (Benjamin & Chun, 2009). Its effectiveness has not yet been validated.

The International Institute for Educational Planning has reviewed current alternative approaches for assessing quality and noted tradeoffs among them (Martin & Stella, 2007). The choices include compulsory versus voluntary quality assurance; universal standards versus assessment of fitness for purpose; minimum versus high standards; and formal accreditation versus more flexible quality standards.

## **D. Concluding remarks**

In deciding on a university admission mechanism, a National Testing Center (NTC) has to be concerned with the tradeoffs among the cost of the examinations, their reliability, and their potential impacts on quality, equity, and corruption.

### **The role of testing and assessment**

A common selection method in high-quality education systems has been to combine low-stakes internal examination results (or course grades) with high-stakes achievement testing. But this approach requires that teachers and administrators are well trained in assessment procedures and that the system is free of corruption. When bribery and cheating are commonplace and teachers and administrators do not have sufficient capacity to assess students, the challenge becomes to develop methods that address short-term needs and at the same time lay the foundation for achieving long-term goals.

A centralized, standardized examination (high-stakes testing) is a relatively inexpensive way to select students to enter university. But the test must be a reliable assessment of skills and knowledge gained through studying the curriculum because the importance of the test results encourages both students and teachers to adopt behaviors that may increase their “success” but reduce their actual achievement. A poorly designed test—one that does not encourage critical thinking or that relies on knowledge unrelated to classroom experience or inaccessible to disadvantaged students—not only harms individual teaching and learning processes, but also lowers educational quality and exacerbates existing inequities across the school system.

High-stakes testing can also be a simple and effective mechanism to deal with corruption, as the recent experiences in Georgia and Ukraine have shown. Public acceptance of the results of high-stakes testing, whether they are used for university selection, teacher evaluation, or broad educational assessment, will be higher when the procedures and instruments are seen as honest and fair. That requires transparency in the development of test content and procedures, control in their application, and objectivity in the scoring and interpretation of results. Publishing standards of behavior and involving civil society, students and parents, in the monitoring process are ways to gain public’s trust. Involvement of teachers in the preparation of instruments has been shown to reduce corruption and it is helpful to computerize as much of the process as possible.

Standardized tests can also promote equity among students of all socioeconomic classes. Although most education systems, like most societies, implicitly tolerate some differences in opportunity and advantage among their members, standardized tests can help to level the field, if two conditions are met. First, the content of the tests cannot be biased against poor or minority students. Second, students from lower-income families must have the means to prepare as well as those from higher-income families. Interventions such as subsidizing low-income students for test preparation time or expense can be one possibility. An

ongoing project in Chile has created a scholarship program with the capacity to provide subsidized test preparation to high-achieving, low-income students. In addition, many countries make special provisions in their admission systems to address the equity issues, for example by lowering cutoff scores for certain disadvantaged groups. In the United States, “affirmative action” programs at many universities have tried various means to promote greater selection of disadvantaged groups in the admission process.

### **The role of NTCs**

A national testing center can both reduce corruption in higher education admissions and improve the quality of secondary education. In order to be effective, the center needs to:

- Have the human and financial resources to design and implement a comprehensive program of assessment. The construction, administration, and scoring of the university selection examinations must be of high technical standard and free of any indication of deviation from regulations.
- Develop a capacity to assist secondary schools in the assessment of the progress of their students. Although the instruments for this assessment may be distinct from the university entrance process, there is strong linkage between the two.
- Play a vital role in the analysis and interpretation of test results, not just for selection into the university, but also for identification of teaching practices and school conditions that enhance learning. The center should be a source of publications designed to inform teachers on the one hand and parents on the other about the how each can best contribute to student learning.

The NTC can also help improve the quality of higher education in the following ways: (i) develop procedures that ensure the competency of students entering university; (ii) assist university faculty in the assessment of their students’ progress and identification of most effective instructional methods; and (iii) develop measures to assess the overall quality of the university in terms of its learning outcomes.

Finally, research on higher education says little about how university admissions processes can improve the level of equity in society. Of course, an honest and fair admission system will reduce inequity in that process of selection. But social inequity is built into the educational system, beginning with preschool education. We cannot expect that the elimination of corruption by itself will have much impact on social equity. In the long term, greater equity will be achieved through fundamental changes in society's principal economic and political institutions. Education's contribution in the short term can be through policies of positive discrimination and equal opportunity. Selection processes can be structured to increase the university attendance of disadvantaged groups, without lowering the overall quality of the university. This task will require not only political will but also careful research. It is a task appropriate for a national testing center.

## References

- Aidt, T. S. (2009). Corruption, institutions, and economic development. *Oxford Review of Economic Policy*, 25 (2), 271-291.
- Aidt, T., Dutta, J., & Sena, V. (2008). Governance regimes, corruption and growth: Theory and evidence. *Journal of Comparative Economics*, 36 (2), 195-220.
- American Educational Research Association. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Amrein, A. L., & Berliner, D. C. (2002). *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*. Arizona State University, Education Policy Research Unit .
- Amundsen, I. (2000). *Corruption: Definitions and concepts*. Bergen, Norway: Chr. Michelsen Institute.
- Anand, P., Mizala, A., & Repetto, A. (2009). Using school scholarships to estimate the effect of private education on the academic achievement of low income students in Chile. *Economics of Education Review*, 28 (3), 370-381.
- Anderson, J., & Photos, I. (2003). *The World Bank and Anti-Corruption in Eastern Europe and Central Asia*. Washington DC: World Bank.
- Anti-Corruption Resource Centre. (2006). *Corruption in the Education Sector*. Bergen: Chr. Michelsen Institute.
- Armstrong, R. D., Jones, D. H., & Wang, Z. (1998). Optimization of classical reliability in test construction. *Journal of Educational and Behavioral Statistics*, 23 (1), 1-17.
- Atkinson, R. C., & Geiser, S. (2009). *Reflections on a century of college admissions tests*. Berkeley CA: Center for Studies in Higher Education, University of California, Berkeley.
- Baker, E. L. (2007). Presidential Address: The end(s) of testing. *Educational Researcher*, 36 (6), 309-317.
- Barclay, P. (2010). *Zimbabwe: Years of hope and despair*. London: Bloomsbury.
- Basturk, R. (2007). Investigation of appointing teachers employed in the public education institutions in Turkey. *Hacettepe Universitesi Egitim Fakultesi Dergisi (H.U. Journal of Education)*, 33, 33-40.
- Barker, D. G. (1967). The history of entrance examinations. *Improving College and University Teaching*, 15 (4), 250-253.
- Beatty, A., Greenwood, M. R., & Linn, R. L. (1999). *Myths and Tradeoffs: The role of tests in undergraduate admissions*. Washington DC: National Research Council ([www.nap.edu/catalog/9632.html](http://www.nap.edu/catalog/9632.html)).
- Benavot, A., & Tanner, E. (2007). *The growth of national learning assessments in the world, 1995-2006*. Paris: UNESCO.
- Benjamin, R., & Chun, M. (2009). *Returning to Learning in an Age of Assessment: A synopsis of the argument. Introducing the rationale of the College Learning Assessment*. New York: Council for Aid to Education.

- Billing, D. (2004). International comparisons and trends in external quality assurance of higher education: Commonality or diversity? *Higher Education*, 47 (2), 113-137.
- Blaug, R. (2010). *How Power Corrupts: Cognition and democracy in organizations*. New York: Palgrave Macmillan.
- Boudett, K. P., Murnane, R. J., City, E., & Moody, L. (2005). Teaching educators how to use student assessment data to improve instruction. *The Phi Delta Kappan*, 86 (9), 700-706.
- Brands, H. (2010). *Crime, Violence and the Crisis in Guatemala*. Carlisle PA: US Army War College, Strategic Studies Institute.
- Braun, S., & Dwenger, N. (2009). Success in the university admission process in Germany: Regional provenance matters. *Higher Education*, 58 (1), 71-80.
- Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, 22, 109-119.
- Brytting, T., Minogue, R., & Morino, V. (2010). *The Anatomy of Fraud and Corruption: Organizational causes and remedies*. Farnham: Gower.
- Burke, R. J., Tomlinson, E. C., & Cooper, C. L. (2010). *Crime and Corruption in Organizations: Why it occurs and what to do about it*. Burlington VT: Ashgate.
- Centre for Higher Education Transformation. (2005). Matric. What is to be Done? *Seminar* (p. 76). Wynberg, South Africa: Centre for Higher Education Transformation ([http://chet.org.za/webfm\\_send/209](http://chet.org.za/webfm_send/209)).
- Chambliss, W. J., Michalowski, R., & Kramer, R. C. (2010). *State Crime in the Global Age*. Cullompton England: Willan Publishers.
- Chun, M. (2002). Looking where the light is better: A review of the literature on assessing higher education quality. *Peer Review*, 4 (2/3), 16-25.
- Cizek, G. J. (1999). *Cheating on Tests: How to do it, detect it, and prevent it*. Mahway NJ: Erlbaum Associates.
- Cohen-Schotanus, J., Muijtens, A. M., Reinders, J., Agsteribbo, J., Van Rossum, H., & Van de Vleuten, C. P. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education*, 40 (10), 1012-1019.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: National Center for Educational Statistics.
- Comité Asesor Técnico. (2008). *Estudio Acerca de la Validez Predictiva de los Factores de Selección a las Universidades del Consejo de Rectores de las Universidades Chilenas*. Santiago, Chile: Consejo de Rectores Universidades Chilenas.
- Contreras, D., Bravo, D., & Sanhueza, C. (2001). PAA, ¿Una prueba de inteligencia? *Perspectivas (Chile)*, 4 (2), 233-249.
- Donoso D., S. (1998). La Reforma Educacional y el sistema de selección de alumnos a las universidades: Impactos y cambios demandados. *Estudios Pedagógicos (Valdivia)*, 24, 19.
- Duhamel, L. (2010). *The KGB Campaign Against Corruption in Moscow*. Pittsburgh PA: University of Pittsburgh Press.
- Evensrel, A. (2010). Institutional and economic determinants of corruption: a cross-section analysis. *Applied Economic Letters*, 17 (6), 551-554.

- Everett, J., Neu, D., & Rahaman, A. S. (2006). The global fight against corruption: A Foucaultian, virtues-ethics framing. *Journal of Business Ethics*, 65 (1), 1-12.
- Eyzaguirre, B. (2003). Exigencias para la construcción de una prueba de selección a la universidad. *Estudios Públicos*, 90 (Otoño), 47-90.
- FairTest. (2006, August). *Exit Exams Decrease Graduation Rates*. Retrieved November 21, 2010, from FairTest: The National Center for Fair and Open Testing: [www.fairtest.org](http://www.fairtest.org)
- Feuer, M. J. (1999). *Uncommon Measures: Equivalence and linkage among educational tests*. Washington DC: National Academy Press.
- Gabrscek, S. (2010). *Comparative Analysis of National Testing Centers in Azerbaijan, Georgia, Ukraine, Kyrgyzstan and recommendations for Tajikistan*. NEPC.
- Gardizi, M., Hussman, K., & Torabi, Y. (2010). *Corrupting the State or State-Crafted Corruption?* Kabul: Afghanistan Research and Evaluation Unit.
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the first year: High school record vs. standardized tests as indicators of four-year college outcomes*. CSFE6.07. Berkeley CA: University of California, Berkeley, Center for Higher Studies in Education.
- Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8 (1), 1-26.
- Goh, D. (2005). NCEA-Bane or boon? *ACEpapers (University of Auckland)* (15), 88-102.
- Gowalski, T., & Lasley, T. J. (2009). *Handbook of Data-Based Decision Making in Education*. New York: Routledge.
- Graff, G., von Maravic, P., & Wagenaar, P. (2010). *The Good Cause: Theoretical perspectives on corruption*. Opladen: Barbara Budrich.
- Greaney, V., & Kellaghan, T. (2008). *Assessing National Achievement Levels in Education*. Washington DC: World Bank.
- Grindle, M. S. (2010). *Good Governance: The inflation of an idea*. Cambridge MA: Harvard University, Kennedy School of Government, Center for International Development.
- Gustafsson, J.-E. (2003). *The predictive ability of grades and aptitude tests in higher education*. Retrieved 11 18, 2010, from Association for Educational Assessment [www.aea-europe.net](http://www.aea-europe.net): [www.aea-europe.net/userfiles/Jan%Eric%Gustafsson%20Keynote%20Presentation.pdf](http://www.aea-europe.net/userfiles/Jan%Eric%Gustafsson%20Keynote%20Presentation.pdf)
- Hallak, J., & Poisson, M. (2007). *Corrupt Schools, Corrupt Universities: What can be done?* Paris: UNESCO: International Institute for Educational Planning.
- Hamilton, S. L., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J., et al. (2009). *Standards-based Accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Corporation.
- Haney, W. M., & Clarke, M. J. (2007). Cheating on Tests: Prevalence, detection and implications for online testing. In E. M. Anderman, & T. B. Murdock, *Psychology of Academic Cheating* (pp. 255-288). Burlington MA: Elsevier Academic Press.
- Haney, W. (2000). The myth of the Texas miracle. *Education Analysis Policy Archives* (<http://epaa.asu.edu/epaa/v8n41>), 8 (41).

- Heilig, J. V., & Darling-Hammond, L. (2008). The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis* , 30 (2), 75-110.
- Helms, R. M. (2009). University Admissions: Practices and Procedures Worldwide. *International Higher Education* , pp. 5-7.
- Heyneman, S. P. (2004). Education and corruption. *International Journal of Educational Development* , 24 (6), 638-648.
- Heyneman, S. P., Anderson, K. H., & Nuraliyeva, N. (2008). The cost of corruption in higher education. *Comparative Education Review* , 52 (1), 1-26.
- Hoover, E. (2008, June). The revised SAT: No better, no worse at its job. *Chronicle of Higher Education* , 54 (42), p. A1.
- Hopkins, K. D. (1998). *Educational and Psychological Measurement and Evaluation. Eighth edition*. Needham Heights MA: Allyn & Bacon.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence of teacher cheating. *Quarterly Journal of Economics* , 118 (3), 843-877.
- Jacob, B. A., Courant, P. N., & Ludwig, J. (2003). Getting inside accountability. *Brookings-Wharton Papers on Urban Affairs* , 41-86.
- Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review* , 54 (1), 5-26.
- Kane, T. J., & Staiger, D. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives* , 16 (4), 91-114.
- Karakaya, I., & Tavsancil, E. (2008). The predictive validity of the university student selection examination. *Educational Sciences: Theory and Practice* , 26 (3-4), 1011-1019.
- Kassahum, D. (2008). Standardization techniques for grade inflation problem at higher educational institutions of Ethiopia: the case of Addis Ababa. *Assessment & Evaluation in Higher Education* , 33 (1), 33-44.
- Kellaghan, T. (2004). *Public Examinations, National and International Assessments, and Educational Policy*. Dublin: St. Patrick's College, Educational Research Centre.
- Kellaghan, T., & Greaney, V. (2001). *Using Assessment to Improve the Quality of Education*. Paris: UNESCO.
- Kerr, K. A., Marsh, J. S., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education* , 112 (4), 496-520.
- Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., & Shavelson, R. (2005). An approach to measuring cognitive outcomes across higher education institutions. *Research in Higher Education* , 46 (3), 251-276.
- Klitgaard, R. (2000). *Subverting Corruption*. Finance and Development (June): 2-5.
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first year-college grade point average. Report No. 2008-5*. New York: The College Board.

- Koljatic, M., & Silva, M. (2006). Validación de la PSU: Comentarios al ' Estudio Acerca de la Validez Predictiva de los Factores de Selección a las Universidad del Consejo de Rectores'. *Estudios Públicos* , 104 (Primavera), 331-346.
- Koljatic, M., Silva, M., & Cofré, R. (2010). Achievement versus aptitude in college admissions: A cautionary note based on evidence from Chile. *The 13th SweSAT Conference. Selection to Higher Education. Fairness, efficiency and consequences* (p. 14). Umea, Sweden: Swedish National Agency for Higher Education.
- Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *The Journal of Human Resources* , 37 (4), 752-777.
- Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education* , 79 (5), 540-563.
- Lewis, M., & Pettersson, G. (2009). *Governance in Education: Raising performance in the sector. Overview of issues and evidence. Draft*. Washington DC: World Bank.
- Libertad y Desarrollo. (2008, December 24). PSU: Truths and myths. 901, pp. 1-4.
- Lissitz, R. W., & Samuelson, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher* , 36 (8), 437-448.
- MacLean, G. E. (1905). The admission of students to college. *Science (New Series)* , 22 (554), 167-175.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield, & M. L. Kornhaber, *Raising Standards or Raising Barriers: Inequality and high-stakes testing in public office* (pp. 85-106). New York: Century Foundation.
- Manzi, J., Bosch, A., Bravo, D., del Pino, G., Donoso, G., & Pizarro, R. (2010). Validez diferencial y sesgo en la predictividad de las pruebas de admisión a las universidades chilenas (PSU). *Revista Iberoamericana de Evaluación Educativa* , 3 (2), 30-46.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Making in Education: Evidence from recent RAND research*. RAND.
- Martin, M., & Stella, A. (2007). *External Quality Assurance in Higher Education: Making Choices*. Paris: UNESCO: International Institute for Educational Planning.
- Mason, P. S. (2010). *The European Union's Fight Against Corruption: The evolving policy toward Member States and candidate countries*. Cambridge England: Cambridge University Press.
- McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior* , 11 (3), 219-232.
- McDonald, A. S. (2001). *Does psychometric testing offer a solution to university selection?* Slough, United Kingdom: National Foundation for Educational Research.
- McDonald, A. S., Newton, P. E., Whetton, C., & Benefield, P. (2001). *Aptitude Testing for University Entrance: A literature review*. Slough, United Kingdom: National Foundation for Educational Research.
- McDonald, M. (2002). The perceived role of diploma examinations in Alberta, Canada. *The Journal of Educational Research* , 96 (1), 21-36.

- McManus, I. C., Smithers, E., Patridge, P., Keeling, A., & Fleming, P. R. (2003). A Levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ*, 327 (7407), 139-142.
- Messick, S. (1989). Validity. In R. L. Linn, *Educational Measurement 3rd Edition* (pp. 13-103). New York: American Council on Education/Macmillan.
- Meyer, L., McClure, J., Walkey, F., McKenzie, L., & Weir, K. (2006, June). *The impact of the NCEA on the student motivation*. Retrieved November 21, 2010, from Ministry of Education, Education Counts: <http://www.educationcounts.govt.nz/>
- Miller, S. (2010). *The Moral Foundations of Social Institutions: A philosophical study*. Cambridge England: Cambridge University Press.
- Mo, P. H. (2001). Corruption and economic growth. *Journal of Comparative Economics*, 29 (1), 66-79.
- Morris, S. D., & Blake, C. H. (2010). *Corruption and Politics in Latin America: National and regional dynamics*. Boulder CO: Lynne Rienner Publishers.
- Moussavi, M., & McGinn, N. (2010). *A Model for Quality of Schooling*. American Associations for the Advancement of Artificial Intelligence. 2010 AAAI Spring Symposium Series.
- National Association for College Admission Counseling. (2008). *Report of the Commission on the Use of Standardized Tests in Undergraduate Admissions*. Arlington, VA: National Association for College Admission Counseling.
- Nelson, L. R. (2006). Using selected indices to monitor cheating on multiple-choice exams. *Journal of Educational Research and Measurement*, 4 (1), 1-18.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral Damage: How high-stakes testing corrupts America's schools*. Cambridge MA: Harvard University Press.
- Niu, S. X., & Tienda, M. (2009). *Testing, Ranking and College Performance: Does high school matter?* Princeton NJ: Princeton University, Office of Population Research ([http://theop.princeton.edu/reports/wp/NiuTienda\\_HighSchoolClassRankandTestScore\\_091809.pdf](http://theop.princeton.edu/reports/wp/NiuTienda_HighSchoolClassRankandTestScore_091809.pdf)).
- Noah, H. J., & Eckstein, M. A. (2001). *Fraud and Education: The worm in the apple*. Lanham MD: Rowman & Littlefield.
- Noeth, R. J., Rutkowski, D. J., & Schleich, B. A. (2010). *College Admission Tests as Measures of High School Accountability*. Bloomington, IN: University of Indiana, Center for Evaluation and Education Policy.
- Nye, J. (1967). Corruption and political development: a cost-benefit analysis. *American Political Science Review*, 61 (2), 417-427.
- Office of Technology Assessment. (1992). *Testing in American Schools: Asking the right questions*. Washington DC: US Government Printing Office.
- Osiptan, A. (2009). Corruption hierarchies in higher education in the former Soviet Bloc. *International Journal of Educational Development*, 29 (3), 322-330.
- Patrinos, H. A., & Kagia, R. (2007). Maximizing the performance of education systems: The case of teacher absenteeism. In J. E. Campos, & S. Pradhan, *The Many Faces of Corruption: Tracking vulnerabilities at the sector level* (pp. 63-88). Washington DC: World Bank.

- Pavez, K. (2008, August<sup>o</sup> 25). PSU: estudio señala que notas de enseñanza media no deberían ponderarse en más de un 25%. *La Tercera*, p. 9.
- Postlethwaite, T. N., & Kellaghan, T. (2008). *National Assessments of Educational Achievement*. Paris: International Institute for Educational Planning.
- Prado Terrazas, S. (2010). Predictive validity of new system admission of Chilean universities: The case of engineering. *The 13th SweSAT Conference. Selectin to Higher Education, Fairness, Efficiency, Consequences* (p. 15). Umea, Sweden: Swedish National Agency for Higher Education.
- Ravitch, D. (2010). *The Life and Death of the Great American School System*. New York: Basic Books.
- Richardson, M., & Abraham, C. (2009). Conscientiousness and achievement motivation predict performance. *European Journal of Personality*, 29 (7), 589-605.
- Robinson, M. (1998). *Corruption and Development*. London: Cass.
- Rose-Ackerman, S. (1978). Corruption. In C. K. Rowley, *The Encyclopedia of Public Choice* (pp. 67-76). Amsterdam: Kluwer Academic Publishers.
- Rosillo López, C. (2010). *La Corruption à la Fin de la République Romaine*. Stuttgart: Steiner.
- Schiefelbein F., E., & McGinn, N. F. (2008). *Learning to Educate: Proposals for the reconstruction of education in Latin America*. Paris: UNESCO/IBE.
- Schiefelbein, E. (1992). Análisis del SIMCE y sugerencias para mejorar su impacto en la calidad. In S. Gómez, *La Realidad en Cifras* (pp. 241-280). Santiago, Chile: FLACSO (Facultad Latinoamericana de Ciencias Sociales).
- Schiefelbein, E., & Schiefelbein, P. (2002). Three decades of testing in Latin America. In C. De Moura Castro, & A. Verdisco, *Making Education Work: Latin American Ideas and Asian Results* (pp. 169-190). Washington DC: Inter-American Development Bank.
- Schmidt-Pfister, D. (2010). *Transnational Advocacy on the Ground: Against corruption in Russia?* Manchester UK: Manchester University Press.
- Schupbach, H., Pixner, J., & Zapf, S. (2006). Handlungskompetenz im Hochschulstudium. *Bildung und Erziehung*, 59 (2), 147-166.
- Shavelson, R. J. (2010). *Measuring College Learning Responsibly: Accountability in a new era*. Stanford CA: Stanford University Press.
- Shaw, P. (2007, November 12). *Educational corruption and growth*. Retrieved November 4, 2010, from [www.econ.yale.edu/seminars/macro/mac09/shaw-090409.pdf](http://www.econ.yale.edu/seminars/macro/mac09/shaw-090409.pdf)
- Shulruf, B., Hattie, J., & Tumen, S. (2008). The predictability of enrolment and first-year university results from secondary school performance: The New Zealand Certificate of Educational Achievement. *Studies in Higher Education*, 33 (6), 685-698.
- Shulruf, B., Hattie, J., Turneraq, R., Tumen, S., & Meisong, L. (2009). Enhancing equal opportunities in higher education: A new merit-based admission model. *Cypriot Journal of Educational Sciences*, 4 (1), 15-32.
- Spector, B. I., Johnston, M., & Dininio, P. (2005). Learning across cases:Trends in anticorruption strategies. In B. I. Spector, *Fighting Corruption in Developing Countries: Strategies and analysis* (pp. 213-232). Bloomfield CT: Kumarian Press.

- Steindl, F. G. (1990). University admission requirements as rent-seeking. *Public Choice* , 65 (3), 274-279.
- Strachan, J. (2002). Assessment in change: Some reflections on the local and international background to the National Certificate of Educational Achievement (NCEA). *New Zealand Annual Review of Education* , 11, 245-273.
- Suen, H. K., & Yu, L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese Civil Service exams. *Comparative Education Review*, 50 (1), 46-65.
- Tanaka, S. (2001). Corruption in education sector development: a suggestion for anticipatory strategy. *International Journal of Educational Management* , 15 (4), 158-166.
- Thorndike, R. M. (2010). *Measurement and Evaluation in Psychology and Education. Eighth edition*. Boston: Pearson.
- Transparency International. (2001). *Global Corruption Report 2001*. Berlin: Transparency International.
- Transparency International. (2005). *Stealing the Future: Corruption in the classroom*. Berlin: Transparency International.
- U4 Anti-Corruption Centre. (2006). *Corruption in the Education Sector*. Bergen, Norway: Chr. Michelsen Institute.
- Underwood, J. (2006). *Digital Technologies and Dishonesty in Examinations and Tests*. Nottingham Trent University. Coventry UK: Qualifications and Curriculum Development Authority.
- Valentino, V. (2007). *Managing Corruption in Higher Education in Moldova*. Cambridge MA: Harvard University, Kennedy School of Government, Master thesis, [http://www.culiuc.com/papers/vvalentino\\_corruption.pdf](http://www.culiuc.com/papers/vvalentino_corruption.pdf).
- Webber, C. (1989). The mandarin mentality: University admissions testing in Europe and Asia. In B. R. Gifford, *Test Policy and the Politics of Opportunity Allocation: The workplace and the law* (pp. 33-57). Amsterdam: Boston MA.
- Wei, S.-J. (n.d.). *Corruption in economic development: Beneficial grease, minor annoyance, or major obstacle?* Retrieved November 4, 2010, from National Bureau of Economic Research: <http://aideffectiveness.zunia.org/uploads/media/knowledge/AidEffectiveness/wei.pdf>
- Weidman, J. C. (2007). *Quantifying the Relationship between Corruption in Education and Economic Development in the Eastern Europe and Eurasia Region: An exploratory literature review*. Washington DC: United States Agency for International Development (USAID/E&E/DGST).
- Weidman, J. C., & Enkhjargal, A. (2008). Corruption in higher education. In D. P. Baker, & A. W. Wiseman, *The Worldwide Transformation of Higher Education* (pp. 63-88). Bingley UK: Emerald Group Publishing.
- Wesolowsky, G. O. (2000). Detecting excessive similarity in answers on multiple choice examinations. *Journal of Applied Statistics* , 27 (7), 909-921.
- Woessmann, L. (2003). *Schooling Resources, Educational Institutions, and School Performance: The International Evidence*. Oxford Bulletin of Economics and Statistics 65(2):117-170.

- World Bank. (2002). *Kazakhstan. Governance and Service Delivery: a diagnostic report*. Washington DC: World Bank, Poverty Reduction and Economic Management Unit, Europe and Central Asia Region.
- World Bank. (2003). *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank and Oxford University Press.
- Yeh, J. P., Herman, J. L., & Rudner, L. M. (1981). *Teachers and Testing: A survey of test use*. Los Angeles CA: University of California, Los Angeles. Center for the Study of Evaluation.
- Zagaris, B. (2010). *International White Collar Crime: Cases and materials*. Cambridge England: Cambridge University Press.
- Zwick, R. (2007). *College Admission Testing*. Arlington VA: National Association for College Admission Counseling.