EVALUATION OF THE LEARNING OF INDIGENOUS STUDENTS IN LATIN AMERICA

Challenges of Measurement and Interpretation in Contexts of Cultural Diversity and Social Inequality[1]

ERNESTO TREVIÑO VILLARREAL

Ernesto Treviño is assistant director in the evaluation of school factors (Evaluación de Factores Escolares), Dirección de Evaluación de Escuelas, Instituto Nacional para la Evaluación de la Educación, Avenida Adolfo López Mateos 437 Oriente, Colonia Centro, CP 20000, Aguascalientes, Aguascalientes., E-MAIL: ernesto.trevino@ineemexico.org

Abstract:
This article analyzes, from the point of view of educational opportunities, the problems of using standardized tests to evaluate the learning of indigenous peoples. By combining concepts of cross-cultural psychology, anthropology and psychometrics, the article presents the principal factors that would affect inferences made from the results of tests given to indigenous students. Lastly, the limitations of standardized tests are summarized, and some recommendations are provided for circumscribing the inferences based on the results of such tests.

Key words: evaluation of learning, intercultural education, equal opportunities, achievement tests, psychometrics.

Introduction

Indigenous peoples have historically been the least served by Latin America's educational systems. In this region of the world, little research had been done on the relation between indigenous groups and educational systems. The indigenous peoples of Latin America, in general, live in conditions of more severe poverty than the non-indigenous population. Figures from 2002 indicate that in Bolivia, three-fourths of the country's indigenous people are poor, compared with one-half of the non-indigenous population. The rate of extreme poverty is 52% of Bolivia's indigenous population, and 27% of the non-indigenous population (Hall and Patrinos, at press). The poverty rates in Ecuador are 84% for the indigenous and 46% for the non-indigenous and the monthly income of indigenous males is only 55% that of non-indigenous men (Winkler, 2004). Poverty indicators vary among the different Maya ethnic groups of Guatemala. For example, the poverty index among different indigenous groups ranges from 63% for the Kaqchikel up to 90% for the Mam.

On the other hand, 42% of the non-indigenous population is classified as poor. The situation is even more unequal on comparing extreme poverty, since the range of the indicator extends from 14% for the Kaqchikel up to 38% for the Q'eqchi, in contrast with 8% for the non-indigenous population (Winkler, 2004). During 2002, 90% of Mexico's indigenous population and 47% of the non-indigenous population lived in conditions of poverty. That same year, the incidence of extreme poverty was 69% for the indigenous and 15% for the non-indigenous (Hall and Patrinos, at press). The figures reveal that indigenous peoples, besides having a high rate of poverty, clearly live at a disadvantage in comparison with the non-indigenous population. This situation determines, to a large degree, the limited opportunities of educational and social mobility for indigenous groups.

All of the educational indicators show that indigenous groups are at a disadvantage. The following figures attempt to illustrate the situation of educational marginalization that affects

indigenous students, without taking into account the type of standardized test. In terms of school attendance in Mexico, 15% of the children between ages six and fourteen do not attend schools in communities where the indigenous population is 70% or more of total inhabitants. This rate is approximately twice the national rate of non-attendance of 7.7% (INEGI, 2000). In Peru, figures from 2000 indicate that 11% of the Quechua population—the nation's largest indigenous group—from ages 6 to 17 does not attend school. In contrast, 9.3% of the Spanish speakers in that age range does not attend school (Kudo, 2002).

Over the long term, school non-attendance translates into unequal years of schooling for the indigenous and non-indigenous. In 1999, the educational level of Ecuador's adult population was 3.5 years more schooling for the non-indigenous. In Guatemala, the average number of years of schooling of the non-indigenous is 5.5 years, more than double the 2.5 years of the indigenous population (Winkler, 2004). In Mexico, a comparison of the two groups in rural areas in 2000, showed that speaking an indigenous language is negatively related to the probability of finishing elementary school, even if the students' socioeconomic characteristics are controlled (Mier y Terán and Rabell, 2003). In recent years, inequalities between the years of schooling of Latin America's indigenous and non-indigenous populations have decreased, however, in part because of the expansion of educational systems in the region America (Mier y Terán and Rabell, 2003; Winkler, 2004). The indicators presented here reveal that educational disadvantages affect Latin America's indigenous peoples, who also have the lowest results on standardized examinations.

The academic achievement levels, shown by standardized examinations, of the indigenous population are significantly lower than those of the non-indigenous population. In Ecuador, fifth-grade indigenous students obtain scores that are 20% lower than those of their non-indigenous schoolmates. The results in Peru are similar, since fourth-grade indigenous students have scores 15% lower than the non-indigenous population (Winkler, 2004). Mexico's indigenous sixth graders in the 2003-2004 school year obtained scores with a standard deviation lower than that of their non-indigenous schoolmates, according to the tests of national standards given by the national institute of evaluating education (Instituto Nacional para la Evaluación de la Educación).

Interpreting indigenous students' results on standardized tests is the central topic of this article. The study attempts to fill a vacuum in Latin America's educational literature by combining the analysis of equity and diversity with the evaluation of the learning of the indigenous population. Most of the literature on indigenous education deals with equity and diversity as two mutually exclusive concepts. These two topics have not been explored in joint form, to consider the tensions and coincidences that exist between them. A dialectic focus on diversity and equity that analyzes the evaluation of learning of indigenous groups permits a better understanding of the dilemmas faced in providing educational opportunities in a context of cultural diversity.

The current project is a theoretical study of the measurement of the learning of indigenous students in Latin America, by means of standardized examinations. The study's objective is to establish the bases for further research on the measurement of student learning in indigenous contexts. This document suggests that inferences based on indigenous peoples' results on standardized examinations, are different from those of non-indigenous populations. The cultural differences between the two groups influence exam results: in the case of indigenous students, test results are not only a measure of the proposed constructs, but also a reflection of a mixture of each culture's specific characteristics and the construct measured. Therefore, the

results obtained by the indigenous on these exams should be interpreted differently than those of other groups. Indigenous students' results must be rated with care, and research must determine the effects of cultural factors and the constructs measured by the exam. Standardized tests could provide information on the aspects of the dominant culture that the indigenous students have learned at school, but the exams' limitations make it difficult for them to offer information on the different forms of learning and knowledge of indigenous cultures. Lastly, since inferences based on exam results could have consequences on equity in education, careful interpretation of such inferences can represent a valuable instrument for promoting the educational opportunities of indigenous boys and girls.

Although studies on educational evaluation through standardized examinations in contexts of diversity have been carried out in various countries, the contrasting cultural, social and economic differences between indigenous groups and the rest of the population in Latin America require a particular specification of the problem. In general, studies of this type have carried out in developed nations that have strong systems of social support for their disadvantaged populations, such as the United States, Canada, Spain, Belgium and Switzerland. In the Latin American context, however, (the most unequal context in the world according to figures from the United Nations), the situation of the indigenous peoples is of extreme marginalization. In addition, the cultural contrasts between indigenous peoples and the dominant population in Latin America are especially acute, since they are based on completely different paradigms of learning. On the other hand, the historical relation between indigenous peoples and dominant cultures has been marked by more than five hundred years of marginalization, exclusion and destruction of indigenous cultures (Falcón, 2002). This relation is qualitatively different from that of the populations in developed nations. Lastly, a premise of the current study is that reflection is required—reflection that considers the specific conditions of the indigenous peoples in Latin America as well as the situation faced by educational evaluation in the region's nations.

The experience accumulated in the field of cross-cultural psychology can contribute to clarifying some of the most controversial topics in the debate on the measurement of educational achievement in non-occidental cultures. As explained below, cross-cultural psychology has been enriched by the incorporation of elements of anthropology that have permitted a more profound understanding of the particular traits of each culture. This experience may aid in preventing the errors committed by the early cross-cultural projects in the field of psychology, especially with respect to the inferences reached by using standardized measuring instruments.

One of the premises of this document is the recognition of the tension between modern and postmodern interpretations of the concept of equity in relation to standardized examinations, in contexts of cultural diversity. On one hand, postmodern focuses believe these examinations discriminate against individuals who do not belong to the dominant culture: the exam serves as an instrument for excluding marginalized cultures while simultaneously legitimating the position of dominant social groups. Within such a framework of logic, equity will be attained when the diverse cultures in a society are equally valued. In a situation of equal cultures, standardized examinations would be meaningless because of their contradiction with the value of diversity. On the other hand, the conception of equity from a modern focus is that standardized examinations play an important role: they measure the skills and knowledge students would need to increase their possibilities of educational and social mobility. This theoretical framework assumes that the social structures that impede the equal valuation of

distinct cultures will not change over the short term; in fact, equal valuation may not even be an objective in this focus. As a consequence, students need to acquire the knowledge and values of the dominant culture, which in this case is modernizing.

The report on this study first presents a classification of standardized tests as a function of various criteria. It then refers to the importance of standardized examinations in contemporary educational systems, international comparisons, and the tensions between standardized evaluation and the growing legal recognition of cultural diversity. The third step in the report is to situate the standardized evaluation that occurs in indigenous contexts, within the field of cross-cultural psychology. In fourth place, the educational practices of indigenous communities and the practices carried out in formal education are compared. The fifth step will be an analysis of the factors that affect the validity of inferences based on standardized evaluation in indigenous contexts, as well as their implications on equity. This section will center its attention on the role played by cultural differences in the validity of certain inferences. In sixth place, questions and proposals regarding equity and evaluation in indigenous contexts will be presented. The end of the document includes general conclusions.

Although this document employs the terms of indigenous culture, indigenous groups, indigenous communities, and indigenous peoples, no suggestion is made that all of these cultures are equal. The terms are used to simplify the explanation. An appeal is made to the reader to consider the broad array of indigenous cultures present in Latin America. In spite of disagreement on the form of counting the indigenous population, Latin America is estimated to have approximately four hundred ethnic and linguistic groups, a reference to the region's cultural heterogeneity (IADB, 2004).

Types of Examinations

Examinations can be classified in two major types, depending on the consequences of the test in the life of students or other educational actors. On one hand, in some systems of evaluation, the results of standardized examinations have few or no consequences (low stakes) for educational actors (students, teachers, or schools). In these cases, evaluation has informative purposes. On the other hand, in other systems of evaluation, test results are used to punish or motivate educational actors (high stakes) (Cizek, 1998; Ravela, 2001). The consequences of the examinations depend directly on the use given to test results. For example, examinations have high consequences in cases that test results are used as selection instruments to admit students in an institution or to credit the successful completion of an educational level. Examinations, even in evaluation systems with no consequences, have an impact on the social appreciation and/or educational opportunities of students.

In the United States, which has had a tradition of examining students to render accounts since the mid 19th century (Hamilton and Koretz, 2002; Kirby and Stecher, 2004), the policies of using high-stakes examinations (revitalized in the late 20th century) are broadening educational inequalities. The effervescence of examinations for rendering accounts had its corollary in the package of reforms known as "No Child Left Behind", which was implemented in December of 2001 to impose on the states the requirement of giving high-stakes tests to students in elementary and secondary school. The logic was that evaluation with consequences would motivate teachers while improving school administration and student learning (Fusarelli, 2004; Kirby and Stecher, 2004).

Studies on the consequences of high-stakes standardized examinations have indicated, however, that the pernicious effects of the model of rendering accounts are detrimental for

students who are poor, immigrants, and/or members of ethnic minorities. Specifically, studies have found a differentiating effect of the policy of rendering accounts, in which students from the middle and high socioeconomic sectors increase their levels of learning. At the same time, the teachers and directors in schools where such students are concentrated, respond to the model of rendering accounts according to expectations; i.e., they are motivated and orient their actions toward improvement. In contrast, students from low socioeconomic levels, immigrants and/or members of ethnic minorities, have shown decreased achievement on examinations, as well as a higher dropout rate and lower levels of schooling. The schools that serve such students tend to implement policies that punish the most disadvantaged students and emphasize the repetitive practice of exercises similar to those included in the standardized examinations, instead of developing the skills that the examinations are supposed to measure (Apple, 2001; Diamond and Spillane, 2004; McNeil, 2000; Muller and Schiller, 2000; Roderick *et al.*, 1999). Based on this evidence, in a hypothetical setting with a system of high-stakes examinations, indigenous students without doubt would be greatly affected by such a measure.

Depending on the consequences of test results, indigenous peoples could encounter at least two situations. In first place, in a setting of measurement with no consequences, indigenous students could be stereotyped as unable to advance successfully through the educational system, since their achievement levels appear to be lower than the non-indigenous population, as mentioned in the introduction. In second place, in settings of examinations with high consequences, indigenous students' results could limit their educational opportunities.

Standardized tests can also be classified by their relation to norms or criteria. Examinations related to norms attempt to compare one student's achievement to other students who have taken the test previously, and whose results have been used to create a scale that serves to categorize achievement in percentiles. Examinations related to criteria attempt to measure the progress of students with functional criteria of performance that have to do with knowledge or skills the students should possess. Such knowledge and skills are reflected in specific test scores (AERA, APA Y NCME, 1999).

Lastly, tests can also be divided into two groups, according to their content. The first group consists of tests related to the curriculum, whose objective is to measure student learning in terms of a predetermined curriculum. The second group corresponds to tests unrelated to the curriculum. These tests are usually designed around criteria established by experts in various fields, who define the skills and/or knowledge students should have.

**Tension between the Growing Importance of
Standardized Examinations and the Recognition of Diversity**

At the present time, the measurement of learning by means of standardized examinations is a common practice that directly influences students' educational opportunities. In some evaluative contexts, examinations have high consequences, since the results are used to make decisions regarding educational policy as well as the financial resources assigned to schools, teachers' salaries and the acceptance or graduation of students at various educational levels. In the United States, for example, the educational reform package called "No Child Left Behind" requires state governments to test all of their students annually, from the first year of elementary school until the last year of secondary education. This reform proposed the use of examinations with high consequences (high stakes) in which schools—and in some cases, students—receive rewards or punishments according to exam results (DOE, 2003). The use of the results of evaluations for rendering accounts gives standardized examinations a leading role

in defining educational opportunities for students: such exams influence the educational future of children and young people because of the extensive effect of exam-based decisions on schools as well as the repertoire of educational options available.

The Asian nations are a classical example of the implications of standardized examinations on the population's future opportunities. In this region of the world, entrance examinations for higher education are a milepost in students' lives. Those who successfully pass the exam will have abundant educational opportunities and greater probabilities to enter a successful field of study (Cheng, 1996). The use of standardized examinations in the Asian context shows the importance the results of such tests have on the population's education and employment possibilities.

In Latin America, entrance examinations also influence educational opportunities. In many countries they determine acceptance into secondary and higher education, and serve as filters that leave a sizeable number of students without an opportunity to access these educational levels. Mexico's institutions of secondary and higher education, for example, have traditionally used entrance examinations to select students. And since the creation of the National Center of Evaluation for Higher Education (Centro Nacional de Evaluación para la Educación Superior—CENEVAL), various public universities in the Mexican states use the same entrance examinations (Tirado *et al.,* 1997). In Chile, students desiring to enroll in higher education must take the University Selection Test (Prueba de Selección Universitaria), which serves as a filter for institutions. Brazil and Colombia also have entrance tests for institutions of higher education. In summary, standardized examinations seem to affect students' educational opportunities primarily when they are used to select those who will enter levels of secondary and higher education.

Entrance examinations could be understood as a pyramidal filter of educational inequality, in which the members of the dominant culture easily attain greater opportunities, while the members of minority cultures are marginalized in the early stages. If educational inequalities are perceived as a series of transition points between levels (Benadusi, 2001), entrance examinations clearly restrict educational mobility or transition to higher educational levels, which are the most valued by a society. Therefore, testing reduces educational opportunities for students with low results on entrance examinations. Up to the present, research in Latin America has not focused on researching the effects of standardized examinations on the educational opportunities of distinct social and cultural groups.

In Latin America, most countries now have areas of educational evaluation to monitor students' achievement, without consequences (Ravela, 2001; UNESCO, 2000; Wolff, 1998). With the exception of entrance examinations for secondary and higher education, the examinations that are used to track schools' performance do not have consequences for the students, teachers, or schools themselves (Treviño, at press).[2] Standardized tests, however, have increased in importance since the early 1990s, and current conditions suggest that their role will be even more important in the future.

International studies on evaluation through standardized examinations explain the relevance of evaluation in multicultural contexts, and in this case, the international setting. The Latin American Laboratory for the Assessment of Education Quality (El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación—LLECE) of UNESCO, the Programme for International Student Assessment (PISA) of OECD, and Trends in International Mathematics and Science Study (TIMSS) of IEA are three outstanding examples of international

projects of evaluation that have aroused the interest of the worldwide educational community. These projects also include a strong component of intercultural comparability.

The experiences of LLECE, PISA and TIMSS offer important insights into the field of evaluation in settings of cultural diversity. In the case of LLECE, the examinations were submitted to a process of consensus among the participating nations, and coinciding curriculum content was identified in the thirteen nations involved in the project. Once the consensus process had concluded, the tests were certified by outside experts (LLECE, 1998). The tests were translated into Portuguese for use in Brazil, which also participated in the consensus process. The PISA study translated and adapted the examinations to ensure that they would measure the constructs they hoped to evaluate in each participating nation. Statistical analyses and consultation with experts were the two main sources of information for adapting the examinations. On completing the cultural adaptation of the tests, the PISA study considered the equivalency of constructs in different cultures (Grisay, 2003; McQueen and Mendelovits, 2003). TIMSS carried out adjustment procedures on the tests to maximize their comparability in the various participating nations; for this purpose, qualified translators were hired and revision processes were carried out in each country (Maxwell, 1996).

The mechanisms to ensure the comparability of results in populations from different countries were not free from criticism. In the case of LLECE, few documents give a detailed account of the proceedings used, for example, in the translation processes. The PISA project, on the other hand, made an exhaustive list of tasks to ensure the cultural comparability of examinations. The PISA study defined the constructs to be measured, developed the instruments and then carried out pilot studies to verify the adequate measurement of the constructs in the participating nations. The selection of constructs, however, was based fundamentally on reading and math skills, with the possible omission of constructs of importance in some of the cultures that participated in the study. The results of TIMSS 1995 show how nations with similar cultures form groups of similar achievement, which seems to be associated with linguistic, geographical, and historical/political factors. Four groups of clearly defined nations exist: English-speaking, German-speaking, Scandinavian, eastern Europe and eastern Asia. The results suggest that these nations share concepts with regard to mathematics and science, as well as the learning and teaching of those subjects (Gronmo, Kjaernsli and Lie, 2004).

These three international studies implicitly assume that cultures are unified within countries—an assumption that is incorrect for most nations. For example, national and international research studies on evaluation usually construct examinations for measuring learning in the official language of the nation in question, without considering that indigenous students have a different native language. Therefore, difference in the mastery of the official language is yet another factor that increases the disparity in evaluation to the detriment of indigenous students.

In brief, international evaluations have developed procedures for adapting and/or translating standardized examinations. These procedures have led to the consideration of aspects ranging from the validity of constructs up to statistical analyses that permit determining the validity of measurement in a certain culture. The absence of non-occidental specifications for different constructs in standardized examinations is probably the most serious error in the adaptation of examinations for these international studies.

Under current conditions, it is easy to foresee that the results of standardized examinations will gradually have greater consequences on students' educational trajectories, and increasingly

influence the educational opportunities of minority groups, especially indigenous peoples. The repercussions of standardized examinations on educational opportunities are not currently considered a priority phenomenon in educational research in Latin America. Thus the use of standardized examinations is thought of as an outside variable, like an unquestionable event that happens in students' lives, in spite of the weight such exams may have on children's educational trajectories.

Standardized examinations become more problematic in the present context, in which nations have progressively recognized the cultural diversity of their populations. For example, twelve countries in Latin America have signed and ratified the Convention concerning Indigenous and Tribal Peoples of the International Labor Organization—Argentina, Bolivia, Brazil, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, Mexico, Paraguay, Peru and Venezuela (ILO, 2003). This convention goes beyond the legal recognition of indigenous peoples by establishing a series of rights including culture, access to land, and social policies.

Situating the Phenomenon in the Field of Cross-cultural Psychology
The use of standardized examinations to measure learning in indigenous communities gives an additional perspective of the tension between etic (view from the outside) and emic (view from the inside) paradigms—tension present in the heart of cross-cultural psychology (Werner, 1979; Suárez-Orozco, 1994). The etic paradigm, related to psychology, seeks to measure and understand certain phenomenon by using standards outside of the context in which the phenomenon occurs. On the contrary, the emic position, more related to anthropology, tries to understand how a certain phenomenon operates in relation to the context and seeks to comprehend the patterns, rites and customs that form the standards of local culture (Werner, 1979). The use of standardized examinations for indigenous peoples is an attempt of the etic paradigm to enter the terrain of the emic, assuming the existence of universal concepts related to knowledge, learning, skills and performance (Van de Vijver, 1997). However, an assumption of the existence of universal constructs in various cultures, as well as the existence of instruments that can measure such constructs, reflects the fundamental contradiction of using the same measure to understand diversity. Such could be the case if the use of the same instrument in different cultures measures phenomena other than the intended construct. Using standardized examinations to understand educational phenomena in culturally diverse populations leaves an unanswered question: What do the results of standardized examinations tell us about achievement in culturally different groups? And more specifically in relation to this research: What is the appropriate interpretation of the results of indigenous students on standardized examinations? At the present time, the answer to both questions would be "We are not sure." The information on learning that is collected through standardized examinations could be reflecting the contrast between the tests' assumptions and local culture, such as differences in learning methods or in the handling of the language used on the exam. Up to now, research has not developed responses to these questions, in part because of the lack of unity among fields of measurement, cross-cultural psychology, and anthropology with regard to the education of indigenous peoples. This text, on articulating these three bodies of literature, seeks to establish the bases for advancing in the theoretical and empirical analysis of measuring educational achievement in indigenous communities.

One of the main challenges of cross-cultural psychology is the problem of categories, which also appears in the measurement of learning in indigenous contexts. This question consists of using a psychological category that is not adequate for the society (Price-Williams, 1975).

Categories may be general or specific in relation to the measurement of learning; general categories are related to the ideology of the societies studied, and may include concepts such as abstract and concrete, and the difference between intellect and emotions (Price-Williams, 1975). These general categories shape the way that distinct cultural groups approach and perceive learning, and are therefore relevant for measuring learning in indigenous communities. On the other hand, specific psychological categories are closely related to cognitive processes and thus directly linked to the measurement of learning through standardized examinations. Some examples of these categories are: the existence of universal linguistic concepts, the presence of mathematical and scientific concepts, the way concepts are created, and the difference between performance and skills (Werner, 1979). The use of psychological categories permits a greater understanding of the limitations and implications of the measurement of educational achievement in indigenous communities by means of instruments that do not necessarily reflect the cultural categories of these groups. Standardized examinations frequently assume the universal existence of a certain type of beliefs, forms of interpreting the world, and processes of teaching and learning, without considering that these categories change from one culture to another (Padilla, 2001). Under these circumstances, the inferences based on the results of a standardized exam given to two culturally different groups are not necessarily equal for both groups. In fact, the assumptions regarding the universality of the concepts measured by examinations limit the inferences and the understanding that the measurement provides about population groups, whose culture is not adequately represented in the constructs that the instrument measures.

The cultural relativism of intelligence is another topic of cross-cultural psychology that is related to the measurement of learning in indigenous contexts. Supporting this concept is the idea that different societies develop distinct "types of intelligence" and use diverse focuses to learn and solve problems (Price-Williams, 1975). When measuring educational learning in indigenous contexts, a contradiction may be found between the type of intelligence that is promoted and cultivated at the local level, and the type of intelligence that is measured through standardized tests. The case may easily be that the type of intelligence that is valued locally and used as a framework for problem solving is different from the type of intelligence the exam measures—in fact, they may be in direct opposition.

The role played by value judgments in the work of cross-cultural psychology is another topic that is related to the measurement of educational achievement in indigenous contexts. Price-Williams (1975) proposes that value judgments play a fundamental role in interpreting the findings of cross-cultural psychology, and that such judgments can lead to conclusions that portray the community under study in a pathological manner, distorting reality. The interaction between value judgments and the use of an inappropriate instrument for a specific population can lead to erroneous conclusions and interpretations. In terms of educational measurement, on interpreting the exam results of students from diverse cultural groups, value judgments can play a negative role by defining these groups as incapable when in reality they are simply not familiar with the tasks assigned on the examination.

As shown, the measurement of the educational achievement of indigenous peoples is a topic that can be situated within a framework of cross-cultural psychology. Therefore, connecting this field with educational measurement would prevent many of the errors committed in the early stages of cross-cultural research.

Knowledge, Learning and Teaching in Indigenous Groups

The purpose of this section is to show the contradictions between focuses for learning in indigenous zones and focuses promoted by formal education. The analysis of such contradictions reveals the discrepancies between the autochthonous educational practices of the indigenous communities and the practices that commonly occur at school. On carrying out this exercise, it is important to take into account that standardized examinations in general attempt to measure constructs related to formal education.

Educational systems promote a type of learning impregnated by occidental notions.[3] Formal education encourages individual work and analytical thinking that follows a linguistic strategy (Tharpe, 1994) in a strict framework of time. Students are expected to learn the content taught exclusively at school, and are evaluated individually.

Within this definition of schooling, an individual must work on his own, becoming involved in analytical tasks and submitting to individual evaluation. In fact, the key mechanism of the legitimization of formal education consists of granting diplomas or certificates, which are conceded only in an individual manner. Formal education also seeks to develop analytical skills, in which comprehension occurs by separating the object of study into different parts. The teaching and learning methods are principally linguistic, and students are frequently a passive agent in learning. Language is the primary vehicle for producing and creating learning. Students learn from verbal explanations, and have few opportunities to learn through practice. At school, students have a limited amount of time for assimilating the content of the curriculum. If a students does not learn the content under these conditions, he is penalized through various mechanisms, including failure.

America's indigenous communities do not necessarily share the Eurocentric vision of education, learning and knowledge. The evidence presented below is based on studies of various indigenous cultures. Tharpe (1994) researched the Navajos and Hawaiian communities, De Haan (1999) analyzed a Mazahua community in Mexico, and Joe (1994) also studied the Navajos. These studies show that indigenous communities differ from the dominant occidental culture in at least five aspects that are relevant to the topic of the measurement of educational achievement. In first place, indigenous cultures have a more collectivist vision of social organization (De Haan, 1999; Tharpe, 1994), a characteristic that also affects the organization of teaching and learning, especially in aspects like pedagogical strategies and the measurement of performance. In additions, variations in the degree of collectivism affect children's cognition and thinking, since the intensity of collectivism in the community is a point of reference that defines the culturally appropriate forms of participating in learning and in educational practices.

In second place, indigenous communities and the dominant culture have distinct conceptions of knowledge. As mentioned above, occidentalized societies promote analytical thought, in which learning occurs when "the whole is revealed through the unfolding of its sections", while native Americans have a holistic perception of thinking in which "the pieces derive their meaning from the tendencies of the whole" (Tharpe, 1994:90). The forms of prevailing thought in each community are closely linked to local education, affecting "perception, problem solving, interpretation and action" (Tharpe, 1994:90). As a consequence, any measurement of educational achievement, in order to be adequate, must consider each community's form of prevailing thought.

In third place, indigenous communities have teaching and learning strategies that are different from those used in occidentalized societies. For example, observational strategies of learning are used extensively, quite different from the linguistics of the dominant culture and

formal education (De Haan, 1999; Tharpe, 1994). In indigenous communities, skill is acquired through practical participation in economic activities in a system that considers students apprentices and gives them sufficient time to learn (De Haan, 1999; Joe, 1994; Tharpe, 1994). Within this system, the timeframe for learning is determined by the needs of the student, who is respected by the person who serves as the teacher. Learning has concrete, symbolic connections with daily life: students learn by participating in specific economic activities related to their immediate context—activities that could become their future occupation (Joe, 1994).

In fourth place, the teaching methods in indigenous contexts differ from those used in formal education. In indigenous communities, learning is a lifelong process that assumes that knowledge and skills cannot be taught; rather, children acquire them through practice (De Haan, 1999; Joe, 1994). In this same sense, teaching/learning methods are generally informal and come from distinct sources (Joe, 1994). In the teaching process, adults assume a non-interventionist attitude, leaving space for error and promoting collective decision-making (De Haan, 1999). Schools, in the traditional sense, are not culturally relevant for indigenous peoples like the Navajo, who do not agree with the idea of forcing children to attend school (Joe, 1994). In summary, teaching methods in indigenous communities are based on learning through practice, with limited intervention by adults and the use of error as a learning opportunity. These characteristics contrast with the principles that reign in formal education.

Lastly, indigenous communities evaluate educational performance based on their own knowledge and educational practices, which are different from those of formal education. Indigenous learning takes place without the possibility of public failure. In such contexts, teachers share the responsibility for error as well as for students' skill levels (De Haan, 1999; Tharpe, 1994). Such practices of evaluation and co-responsibility in learning are contrary to the intentions of standardized examinations created to evaluate students in an individual manner.

In summary, the cultural characteristics of indigenous communities, especially those related to learning, are powerful factors that can compromise the validity of inferences based on the results of standardized examinations: tests do not necessarily take into consideration the local forms of learning and measurement. In essence, examinations attempt to measure constructs relative to the learning, knowledge and skills that could be assumed equal in different cultures. Yet, although these constructs exist in different cultural groups, their qualities vary according to the local culture: different groups have created culturally relevant conceptions of knowledge, learning and teaching.

Validity, Culture and Equity
when Measuring the Learning of Indigenous Peoples
This section will analyze the areas of validity, culture and equity, which must be considered when studying educational achievement in indigenous cultures. Before beginning this analysis, it is important to emphasize certain points that will aid in providing a framework for the discussion.

The first aspect to take into account is that standardized examinations commonly measure skills related to formal education, even if they do not measure the contents of the official curriculum. Because of this characteristic, the content of examinations can be considered to be an attribute that is determined externally—in this case by the educational authorities—and over which those in charge of developing the examinations have no control. In this sense, content is not necessarily a problem of the test. On the other hand, examinations have intrinsic

technical characteristics over which those in charge of developing the exams have greater control, such as format, discrimination of questions, and reliability.

The second point of importance for this discussion is the improper use of test results. The argument can be made that such use is a generalized practice, especially with regard to incorrect inferences in relation to culturally diverse populations. Unfortunately, on repeated occasions, inferences are not necessarily supported by the tests' technical characteristics and purposes. Thus we frequently find that examinations are used as a generic product to measure educational achievement, without considering the technical differences of tests. In this context, concepts like validity, bias and adverse impact are rarely considered when making inferences.

In accordance with the above idea, a third point of importance is the distinction among validity, the exam's bias, and adverse impact. Validity is a property of inference based on exam results. An inference is valid if it is aligned with the purposes of the exam, assuming that the test correctly measures the constructs of interest. In second place, bias is a statistical property of the test that could affect validity. Statistical evidence of the existence of bias, however, must be accompanied by substantive evidence on the exam's form and content. On analyzing a test's bias, it is important to consider that different results from various cultural groups are necessary but insufficient evidence to ensure that the exam is biased. The third concept, regularly confused in discussions of validity and bias, is that of adverse impact. According to this concept, an exam may not show a bias yet may consistently affect students who are members of low-performing groups (Camilli and Shepard, 1994). Therefore, the use of examinations in decision-making could often affect these groups disproportionately, since the test results are used erroneously for making decisions.

Factors that Affect Validity
According to the Standards for Educational and Psychological Testing, "validity refers to the degree evidence and theory support the interpretation of test results in terms of their proposed use" (AERA, APA and NCME, 1999:9; Messick, 1993). The inferences derived from test results must be supported by evidence of their validity. Different types of evidence support the validity of inferences, including: exam content, the response process, the test's internal structure, the relation between test results and other variables, the validity of the construct and the consequences of the tests (AERA, APA and ncme, 1999). These types of evidence of the validity of inferences are fundamental for analyzing the educational achievement of indigenous students. First, as shown above, the content of standardized examinations does not necessarily reflect the type of knowledge, learning and teaching methods that are prevalent in indigenous communities. This lack of agreement can by characterized as a typical problem of categorization in the field of cross-cultural psychology—the problem of using an inadequate psychological category for the group under study (Price-Williams, 1975). In this case, standardized examinations could be inappropriate for measuring constructs related to education in indigenous groups, since they differ epistemologically from the constructs promoted by formal education. As the above paragraphs show, the types of knowledge, learning, teaching, and methods of evaluation used by indigenous communities are habitually in opposition to the conceptions present in formal education and standardized evaluation. For example, test content could be based on analytical focuses that are not necessarily shared by indigenous communities; on the other hand, exams could be attempting to measure skills that are not valued or encouraged in the indigenous setting. Such differences between indigenous

and non-indigenous cultures become factors that affect the validity of inferences based on the results obtained by indigenous students on standardized examinations.

Bias is a characteristic of the internal structure of examinations that can also invalidate inferences based on measurements in culturally diverse contexts. Bias in examinations can be conceived as a "systematic difference in the validity of the inference, so that it is less valid for one group in comparison with another" (Koretz, 2003, personal communication). Camilli and Shepard (1994) define bias as "invalidity or systematic error in an exam's measurement of members of a particular group" (p. 8). According to these definitions, bias is a statistical property of examinations in which systematic error in measurement affects a group's members. An analysis of bias, however, requires paying attention to the statistical properties and substantive characteristics of the test, as well as the inferences that can be made (Camilli and Shepard, 1994; Popham, 2000). It should be emphasized that differences in test scores are necessary but insufficient evidence of bias in examinations. On the other hand, bias in standardized tests could affect indigenous peoples, since this type of examinations measures educational constructs that are qualitatively different from those valued by indigenous groups. Examinations may also include common situations and examples from the dominant culture that may appear out of context for indigenous cultures.

International evaluation projects through standardized examinations like TIMSS and PISA have developed processes to decrease the bias of examinations for their use in different national contexts. Although, as stated previously, these initiatives consider countries as homogenous cultures, the processes to minimize bias can illuminate the field of evaluation in contexts of ethnic and linguistic diversity. TIMSS and PISA made a careful selection of questions, considering cultural bias and discarding questions that would require basic cultural knowledge for their correct interpretation. For example, in the case of TIMSS, questions were discarded that implicitly required students to have a knowledge of baseball, since many of the participating cultures are unfamiliar with the sport (Maxwell, 1996).

In the case of the PISA reading tests, steps were taken to minimize the effect of cultural differences on the results. During the tests, special attention was paid to the equivalence of constructs, exam format, and speed in answering. To ensure equivalency, the constructs were defined in direct consultation with experts representing each country. In addition, the PISA project considered different text formats and contexts that could be familiar for fifteen-year-olds in various countries. Another concern was using a varied range of stimuli in the test materials (McQueen and Mendelovits, 2003). For test translation, PISA implemented a strategy that required each country to translate into its official language the two original versions of the examinations, which were written in English and French. Thus the test was created in each country's official language. Lastly, a team of verifiers selected by PISA compared the original versions with those produced by each country (Grisay, 2003). In addition to the translation, the examinations were adapted by substituting terms, modifying syntax and changing the names of people, places, and organizations, when the adaptation would improve the linguistic and cultural equivalency, without affecting the difficulty of the texts' stimuli. Once the versions for each country were ready, pilot studies were completed and the questions were analyzed in each culture (McQueen and Mendelovits, 2003).

As international experiences show, the bias of standardized examinations can be reduced by means of a broad, deliberate and technically adequate effort. Even in circumstances in which test writers take steps to minimize the bias, it dos not necessarily disappear. Therefore, bias

must be taken into account when basing inferences on the results of standardized tests in diverse contexts.
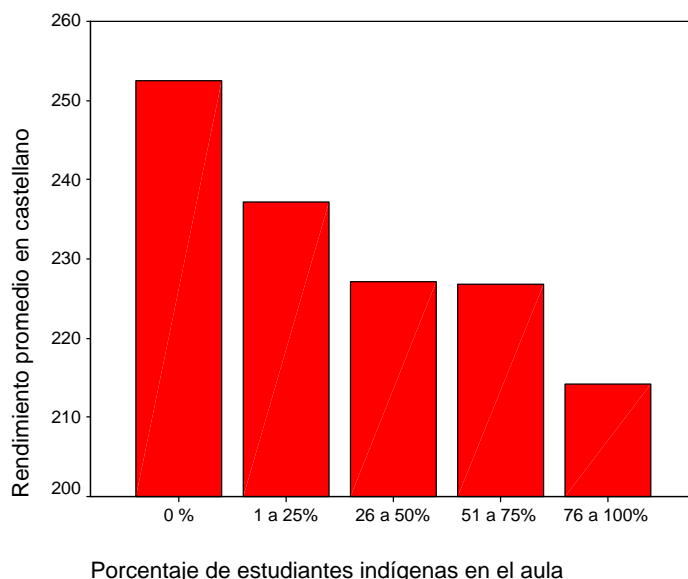
Evidence on the validity of inferences can arise from the relation between the results from standardized examinations and other variables (AERA, APA y NCME, 1999). In the case of indigenous communities, it is important to analyze the relation between belonging to a group and test scores: this relation may help to clarify either the meaning of the results for each cultural group, or the difference between construct measurement in indigenous and non-indigenous peoples.

Graph 1 shows the average scores on the language test of Latin American Laboratory of the Evaluation of Quality of Education (Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación) for classrooms with different percentages of indigenous students in Bolivia, Mexico and Peru. These test scores have been standardized on a scale having a mean of 250 and a standard deviation of 50. As shown in Graph 1, as the percentage of indigenous children in the classroom increases, the average scores decrease, with a difference of 3/4 of standard deviation between the scores in classrooms without indigenous children and classrooms with 75% indigenous students.

The results of Graph 1 show that the scores on the language test are different in classrooms whose racial composition is different; disparities in scores could indicate that "the meaning of the scores is different for members of the different groups" due to the under-representation or irrelevance of constructs in the different population groups (AERA, APA and NCME, 1999:15). These results suggest the plausibility of the hypothesis that the constructs the tests are attempting to measure are inadequately specified for their use in indigenous contexts. If this hypothesis is proven, the first obvious question that comes to mind, in terms of the representation of constructs, is the linguistic incompatibility between the language used on the exam and the native language of indigenous students. Thus indigenous students would be at a disadvantage in comparison with native speakers of Spanish. As a result, scores would have a different interpretation for each group: for Spanish speakers, they would reflect the linguistic skills measured by the exam, such as reading comprehension skills and the expression of ideas through writing; for the indigenous students, test scores would represent their level of handling their second language, in this case Spanish. As a consequence, possible inferences of the results of the examinations would be different for the groups under study.

GRAPH 1
Average Performance of Students in Language, by the Percentage of Indigenous Children in Third- and Fourth-Grade Classrooms, for Bolivia, Mexico and Peru

Porcentaje de estudiantes indígenas en el aula

The above discussion reveals the importance of the notion of construct validity, which refers to the degree that test results reflect student performance on the concept the test is attempting to measure. The presented example suggests that test results should be questioned from the optic of contextual specificity: we must ask if the way the construct is measured is specific to a context or if, on the contrary, it can be generalized to various contexts (Messick, 1993). In sum, the generalization of the underlying constructs of an exam must be analyzed when the tests are used in contexts of diversity.

The validity of the consequences of examinations is another source of evidence for judging inferences. Here it is important to distinguish between the consequences of examinations as evidence to judge the validity of inferences, and the consequences of policy decisions based on test results (AERA, APA and NCME, 1999). In the case of the educational measurement of indigenous peoples, discriminatory interpretations and bias in attributing responsibility for exam results could have two types of consequences. First, the inappropriate use of tests could lead to interpretations of academic performance that portray indigenous peoples as incapable of obtaining success at school, based on the results of examinations that are not even written in the students' native language. In this area, examination scores could be used to support racist theses, to explain the different performance of distinct cultural groups. The second factor that would affect validity would be the causal attribution of the results. In this sense, it is necessary to develop multivariate analyses with complex research designs, in order to determine the causes behind the students' results; the purpose of the examinations themselves is to describe rather than explain a situation.

On the other hand, if inferences are limited correctly by using the appropriate analytical methods, the results of such analyses will permit establishing causal relations between the different scholastic and extra-scholastic factors associated with student performance. These consequences are not always considered in analyses of the validity of evaluation instruments used for culturally diverse populations. The interpretation of the scores might therefore be based on a pathological framework, which would portray indigenous students as lacking in the abilities valued by the dominant culture, but present in a different form in their own culture. Although theorists attempt to distinguish between bias and consequences, and emphasize that

an exam may have an adverse impact on low-performing groups without being biased, the truth is that the low performance of a group, independent from the bias of the exam, may have an impact on validity in relation to consequences.

The Role of Culture in the Validity and Interpretation of Examinations
The cultural validity of standardized examinations is a type of evidence commonly not referred to in the field of educational measurement. Cultural validity is crucial for understanding the results obtained by indigenous students on standardized examinations. Cultural differences may be a factor that affects the inferences made from exam results, because exams may not adequately measure the types of knowledge, learning and skills of a particular group of students. Solano-Flores and Nelson-Barber (2001) propose the concept of cultural validity specifically in science examinations, suggesting that students' sociocultural characteristics influence their form of thinking as well as their understanding of science. Such "sociocultural influences include the values, beliefs, experiences, patterns of communication, styles of teaching and learning, and epistemologies inherent to students' cultural origin and the prevailing socioeconomic conditions in their group" (Solano-Flores and Nelson-Barber, 2001:555). Indigenous peoples have different forms of understanding the world and the ideology promoted by formal education. Such differences must be taken into account when evaluating students that belong to distinct cultural groups and when making inferences based on the results they obtain on examinations. The concept of cultural validity suggests that to design examinations and analyze the results, limits will need to be placed on the affirmation that universal concepts or constructs exist in different cultures.

On occasions, the factors that affect the validity of examinations are confused with bias. For example, Padilla (2001) affirms that the three main factors that bias standardized examinations are:

1) The bias of content, which gives an unfair advantage to the group best represented in the exam's content.
2) The characteristics of evaluation, such as the test format and administrative proceedings that could favor one group to the detriment of others.
3) The inappropriate use of examinations as predictors of future academic performance.

Of the three factors mentioned by Padilla, only exam content can be considered a source of bias. However, the conclusion that an evaluation instrument is biased requires, in addition to statistical evidence, a substantive analysis of content. The characteristics of the evaluation are factors that could affect validity when measuring academic performance in culturally diverse populations, yet they cannot necessarily be considered as bias. Lastly, the erroneous use of tests does not represent a source of bias since "use" is not an exam property. In spite of this confusion, the three characteristics mentioned by Padilla are factors that could affect the cultural validity of examinations. For example, some indigenous groups maintain a sense of shared responsibility for performance, which contradicts the individualized form of measuring educational results on standardized examinations. The inappropriate use of tests is another factor that could affect exam validity because of the consequences on students' schooling, especially when the results are used for making decisions.

The cultural validity of examinations is especially important in today's multicultural societies, since the concept is fundamental for understanding the limitations of standardized

tests for students from distinct cultural groups, as well as for establishing the limits of inferences that can be based on information provided by examinations (Solano-Flores and Nelson-Barber, 2001). Studies of cultural validity will help rate the findings related to test scores, and will offer broader comprehension of educational results in relation to the context.

The technical definitions of validity, bias and adverse impact do not solve other problems with examinations, such as the supposition of the universality of valid constructs for different cultures and the inappropriate use of results for making inferences (Hambleton, 1993; Messick, 1993; Van de Vijver, 1997). It is clear that constructs are not necessarily universal, since different cultures have developed distinct types of knowledge, teaching and learning. The epistemological orientation of diverse cultures could cause the same construct to operate in a different form in various groups. For example, in the Aymara culture, if a teacher makes the sign shown in Drawing 1 and asks his students, "How many fingers do I have here?" his students will answer, if they know how to count, "two fingers". If the teacher changes the gesture and shows his students the sign shown below (Drawing 2) and asks, "And how many fingers do I have now? the Aymara students will answer "You are missing one", instead of saying, "You have one".

Drawing 1

Drawing 2

Does this example mean that Aymara students lack the knowledge to do mathematical subtractions? Certainly not, because they were able to identify the essence of the remainder; yet they used an holistic frame of reference. On standardized examinations, the answer given by the Aymara children would probably be considered incorrect, because this alternative would not necessarily be included as one of the multiple choices for an answer. The example shows the distance between the epistemologies of different cultural groups. Although Aymara children are able to complete the subtraction process that underlies the question, the answer they offer from their interpretative framework would be inadequate from an analytical perspective. If the question were included on an exam given to Aymara students, and the only accepted right answer were "you have one finger", the affirmation could be made that the test suffers from the under-representation of the construct it is attempting to measure. This is because it excludes important dimensions from the construct (Messick, 1993)—dimensions that are specific to the indigenous context. Therefore, the test would be ineffective in measuring the mastery of subtraction among Aymara children. In short, the inferences based on the results of this hypothetical test could not be generalized to the indigenous context (Messick, 1993).

Standardized examinations may be unable to measure certain skills in different cultures. As shown above, indigenous cultures encourage knowledge and skills different from those learned

in formal education. For example, Cole (1996) found different strategies for categorizing objects when comparing the Kpelle of Liberia with Americans and Canadians. This research discovered that variations in the method used to measure the classification of objects was an important influence on the Kpelle's performance. On the other hand, the Americans and Canadians focused principally on looking for paired categories, although they did not necessarily exist in the experiment. Similarly, research in Yucatan, Mexico, found that indigenous as well as mestizo individuals with more schooling had a higher performance on carrying out theoretical activities and solving simple syllogisms (Cole, 1996). This result suggests that people with more education tend to perform better on activities of a scholastic type, although they may be able to apply the same skill in daily life in a different manner from the way it is taught in school.

The validity of inferences based on standardized examinations may also be affected by the assumptions of cultural comparability underlying the tests. Such examinations have three types of fallacies with respect to cultural comparability: *a)* contextual equivalence—the children taking the exam are familiar with the norms of discourse used; *b)* conceptual equivalence—the students are familiar with the attributes of the tasks included on the exam; and *c)* linguistic equivalence—the language used in the evaluation has the same meaning for children of different cultural origins (Gopaul-McNicol and Armour-Thomas, 2002). Contextual, conceptual and linguistic equivalence are directly related to the assumption of the universality of constructs. Standardized examinations may be discriminatory because they are biased toward the dominant culture (Garcia and Pearson, 1994, quoted by NCR, 1997). This type of bias has political implications that go beyond the technical aspects of the examinations, given the questioning of the process through which certain types of knowledge are considered more important than others. In general, standardized examinations do not incorporate the experiences of minority groups (Padilla, 2001). Such questioning enters the field of equity and tensions among diversity, standardized examinations, and the use of exam results.

On measuring student achievement and making inferences based on exam results, it is key to consider the role of language, previous experiences and the lack of learning opportunities. Students from different cultural and social groups experience school in different ways, by dealing with complex relations with factors such as the hegemonic values of the dominant culture, the resistance of the belief systems of marginalized cultures, the incompatibility of the cultures of school and home, and the difficulties of acculturation (Gopaul-McNicol and Armour-Thomas, 2002). These differences, along with disparities among schools, affect the educational achievement of students from diverse cultural groups.

Current Perspectives in the
Cultural Adaptation of Standardized Examinations

At the present time, various groups are developing cultural adaptations of standardized examinations. The international projects of educational evaluation are the most outstanding example of this adaptation, although many countries are adapting standardized examinations on their own. In the international setting, the contributions of Hambleton have the greatest influence, since they were used as the basis for the TIMSS and PISA examination. In a seminal project on international comparison, Hambleton (1993) proposed methods to ensure the adequate translation of examinations into various languages, and underlined the importance of appropriate cultural adaptations. According to this author, key aspects must be considered in the cultural adaptation of standardized tests. The levels of collectivism/individualism of

cultures, as previously mentioned, affect the way their members confront individual evaluations. Familiarity with the exam format, motivation and exam anxiety are other variables that can affect the validity of inferences based on the results of standardized tests in different cultures (Hambleton, 1993). In a later project, Hambleton and Pastula (2000) proposed the existence of three sources of error in the cultural adaptation of examinations: cultural and linguistic differences, techniques of adaptation, and the interpretation of results.

The cultural adaptation of standardized examinations has been analyzed from various paradigms, with the use of statistical methods that attempt to determine the effect of diverse variables on exam results. A meta-analysis of the first twenty-five volumes of the *Journal of Cross-Cultural Psychology* found that the hypothesis of cultural bias in examinations best predicted the performance of different cultural groups. This finding confirmed that exam content is biased toward occidental culture, and that tests do not taken into account the measurement of knowledge and skills unrelated to non-occidental cultures (Van de Vijver, 1997). On the other hand, verbal/cultural aspects have also been found to influence the results of standardized examinations in comparison with the tests' cognitive complexity—a finding obtained by contrasting the results of immigrants and non-immigrants in the Netherlands (Helms-Lorenz, Van de Vijver and Poortinga, 2003). In short, substantial evidence supports the hypothesis that examinations have cultural bias that presents challenges when making inferences based on the performance of students belonging to different cultural groups.

The results of these projects are revealing for the field of measurement, but they suffer from some problems. In first place, statistical analyses may show a bias in selection, since most of the results deal primarily with tasks typical of the formal education of occidental culture. Therefore, the sample of the type of tasks included in these tests is not representative of the universe of tasks that exist in the various cultural groups evaluated (Van de Vijver, 1997). In second place, the lack of relevant tasks for non-occidental cultures on standardized tests presents two problems: on one hand, we do not know if students from non-occidental cultures really handle the knowledge or skills that the examinations attempt to measure, since questions that reflect their modes of learning are not included; and on the other hand, evaluation by means of standardized examinations may clash with the learning theories that assume the existence of different ways of learning a single concept (Bransford, Brown and Cocking, 2000). In conclusion, evaluations must be made more flexible to include tasks relative to non-occidental cultures, and thus study in greater detail the ways performance can be evaluated among students in indigenous contexts.

Equity and Diversity
Evaluation through standardized examinations is a example of continual tension between diversity and equity. On one hand, the recognition of cultural diversity has grown within nations, with effects on the evaluation of language in some contexts (Kalantzis *et al.,* 1989). However, the recognition of diversity does not seem to have had an effect on the implementation of exams' underlying constructs. In the Latin American context, only Bolivia has given tests written in indigenous languages with culturally relevant constructs (Treviño, at press). Yet the recognition of diversity within national borders is far from influencing the operation of national systems of evaluation, which are mainly oriented to measuring the learning of curriculum content, without necessarily incorporating knowledge from the various cultures that live in each country.

From the viewpoint of educational evaluation, the recognition of diversity could oppose some aspects of equity. Even if evaluation systems were to adapt to nations' linguistic and cultural diversity, the prevailing social structures would impose on indigenous peoples the need to learn the dominant language to increase their possibilities of advancing in the educational system, accessing social services, and strengthening social equity (Kalantzis *et al.,* 1989). This tension is typical of the systems of Latin America, where multiculturalism has been recognized, but the indigenous must handle the dominant culture to achieve social and economic mobility. In these conditions, evaluation systems face powerful incentives to measure the aspects that social and economic structures expect to correspond exactly to the attributes promoted by formal education and molded by the dominant culture.

No definite solution exists for this paradox beyond the possibility of attaining an intermediate agreement that respects diversity while monitoring equity. Such an agreement requires that any inference based on the results of standardized examinations take into account the cultural, linguistic, and social characteristics of the evaluated population. On the other hand, as long as social structures make the handling of the dominant culture's codes a requirement for social mobility, it will be necessary to monitor students' performance as a way of evaluating the performance of the educational system in terms of equity.

The consequences of evaluation are a crucial factor in the tension between equity and diversity. The use of standardized examinations in making decisions could have significant impact on children's educational opportunities, especially children belonging to indigenous cultures. Therefore, standardized examinations should not be the only evaluation instrument for judging an individual's academic performance, nor should results be held as a reliable forecast of future academic performance, for two reasons: first, standardized examinations are unable to measure all contents learned at school; and second, the reliability of results decreases on comparing a sample of the evaluated population, and assigning greater probability of error to the estimate of individual learning. Standardized examinations must be used in an adequate manner, with respect for their limitations, and they should not be used to restrict children's educational opportunities.

Equity and Measurement in Indigenous Contexts
The immense educational inequalities of Latin America and the conditions of marginalization in which indigenous groups live will not be solved simply through adequate evaluation. Actions and policies will be required, beyond the scope of this document.

From the perspective of equity, it is essential to question if evaluation through standardized examinations, with all their limitations and consequences, is worthwhile in contexts of cultural diversity and enormous educational, social and economic inequalities. In first place, the use of standardized examinations has been fertile ground for debates of an epistemological, pedagogical and ideological type, with regard to the constitution of legitimate knowledge and educational inequalities; standardized examinations implicitly legitimate certain knowledge as valid and valuable. In this sense, the diverse organizations that use standardized tests sanction that knowledge as valuable and discard other types of knowledge that are excluded from exam design. Social groups that have knowledge different from legitimated knowledge are therefore marginalized for this difference, with consequences that can range from the lack of visibility of these groups' epistemological traditions, up to the loss of educational opportunities and social mobility due to having failed an exam. In pedagogical terms, the use of high-stakes examinations limits the vision of education to the acquisition of the knowledge and skills that

are measured by the examinations. Therefore, the issuing of evaluative judgments on education is based primarily on exam results, and ignores important aspects in which schools and teachers are an influence, such as values, learning for interaction, respect for others, and many other areas where education plays a fundamental role in the socialization of new generations. In terms of ideology, examinations face the paradigms of modernity and post-modernity. Modernity, taken in its extreme version, has a linear vision of economic progress and development, in which societies will move through various stages on the way to industrialization and post-industrialization. According to this focus, the use of standardized examinations is legitimate. Assuming that examinations tend to focus on analytical capacities, the viewpoint of modernity does not perceive as problematic the measurement of only knowledge and skills of an analytical nature, which are appreciated as valuable due to societies' movement through various stages of development. On the other hand, post-modernity, also in its most radical form, considers that all types of knowledge are legitimate and that no type should prevail over another. This focus completely rejects the use of examinations as elements for legitimating the knowledge and power of one group over another. Both postures have critics who consider standardized examinations to be an institutional mechanism for marginalizing disadvantaged students and legitimating the favored position of dominant groups. The evidence shown in this document seems to support the theory that examinations of high impact are perverse mechanisms that widen the learning gaps between socioeconomic and cultural groups.

Standardized evaluations suffer from serious methodological limitations for their use in contexts of diversity. Even in countries with long traditions of standardized evaluation, scientific progress in adapting examinations for diversity is in an initial stage of theoretical and methodological conception. Statistical analyses to evaluate the presence of bias and adverse impact are unable to solve the underlying problem of the epistemological, more than linguistic, disparity among diverse cultures.

The possibilities of making the adequate adaptations and interpretations of exam results in Latin America are also limited, given the lack of specialists in the field and the topic's low priority on research agendas.

The discussion of equity suggests that evaluation in contexts of diversity can lead to serious consequences for marginalized groups—in this case, indigenous groups—and should therefore be subject to severe judgment. Greater cultural and social marginalization, as well as fewer educational opportunities, are consequences associated with the use of standardized examinations in culturally diverse groups. On the other hand, from the scientific point of view, it is difficult to rate the inferences that can be based on the results of standardized examinations.

Most evaluations of indigenous students in Latin America have completely discarded the relation between cultural diversity and results on standardized examinations. Although many evaluation systems in the region are relatively young, and are in the process of consolidation, their newness does not justify the installation of simplistic views of the standardized evaluation of indigenous peoples that discredit their knowledge. In cases of high-stakes testing, such evaluation may have direct negative effects on students' educational opportunities.

A unavoidable fact is that evaluation through standardized testing is already part of the agenda of Latin America's educational authorities. Faced with this reality, the limitations of this type of evaluation in indigenous contexts have already been stated, as well as its consequences for equity. The following section, in light of the imminent presence of standardized evaluation

in the region, offers some simple recommendations that could increase our understanding of evaluation in contexts of diversity, while reducing the adverse, unfair impact of such examinations when given to indigenous students.

Recommendations for a Fair Evaluation of Indigenous Students

Having listed the possible difficulties and perverse effects of standardized evaluation in indigenous contexts, this article will now mention some recommendations to minimize the adverse impact and to advance in the field of knowledge. This section will discuss some of the challenges of educational evaluation in indigenous contexts in Latin America; it will address ways the exam results of indigenous students can be appropriately used. It will also present strategies to aid in improving standardized examinations. Lastly, it will touch on the research of educational measurement in the multicultural contexts of Latin America.

Latin America seems to have the technical ability to make appropriate evaluations of indigenous cultures, but the possibilities of sustaining this type of evaluation over time are weak. The available evidence indicates that Bolivia has done evaluations with examinations adapted to indigenous cultures and languages (Barrera, 1998). Thus the evaluation of intercultural educational is technically feasible, although the quality of these tests should be corroborated. Although the areas of evaluation in Latin American nations have the human ability for the technical development of intercultural evaluations, the Bolivian experience shows that this type of measurement has been unsustainable over time. Therefore, political and financial support for this type of evaluations will probably be weak.

Under current conditions, we are unfamiliar with the meaning of the scores obtained by indigenous students on various tests. Indigenous children take standardized evaluations written in Spanish (or in Portuguese in Brazil), and the results are commonly interpreted as irrefutable proof of their levels of learning. We lack analyses, however, that will permit evaluating the function of examinations in different cultures, while adequately rating the inferences based on the results of indigenous students.

We are also unfamiliar with the degree of validity of the tests' constructs in indigenous contexts. Thus indigenous epistemological traditions may generate valid answers in the various areas that examinations attempt to measure, yet such answers may not necessarily be considered correct on tests based on epistemological currents different from those of indigenous groups.

The combination of factors such as social inequalities, encouragement for high-stakes evaluation, and unfamiliarity with the significance of indigenous children's results on standardized tests could have serious aftereffects on the opportunities of the indigenous. In a context that refers increasingly to the rendering of accounts, it would be hazardous to forget that we live in the most unequal region of the world, where high-stakes examinations would cause even further damage to historically disadvantaged populations, like indigenous groups. In such a scenario, evaluations would not simply ratify the disadvantageous condition of indigenous students' handling of the symbols of formal education, but would also create an unfair barrier for indigenous students who live in conditions of marginalization.

Standardized evaluations have a series of limitations that must be considered when using them for indigenous students. A simple step in this direction refers to the correct interpretation of results, along with a clear delimitation of inferences. Factors such as the exam's purposes, content, language and format, are important characteristics that could affect the validity of inferences. Thus the scope of inferences must be established. If bias and other

factors that affect the validity of inferences are evaluated carefully, reasonable, sustainable conclusions can be reached to help describe the performance of indigenous students on the various constructs that the exam is attempting to measure. This exercise must take into account that examination scores describe a situation and not the reasons behind the various levels of learning.

Considering the equity of educational opportunities and assuming that examinations measure their underlying constructs adequately, the factors that affect validity could be maintained within acceptable margins, and valuable, adequate inferences based on exam results could be reached. For example, since most educational systems in Latin America do not offer options for bilingual or multicultural education beyond elementary school, the scores of tests that use the language of the dominant culture would show the achievement levels of indigenous students in terms of the tasks of formal education. Based on this information, the inference could be made that the low performance of indigenous students could limit their future educational opportunities, especially in their transition to more advanced educational levels. In this form, test results would offer important information on the present inequalities that influence future educational opportunities for indigenous students. This valuable information can be obtained by using available data on student performance. The only requirement, not problem-free, is that students' ethnic affiliation be identified on each exam.

Another relatively simple measure for improving examinations would be a sampling of various types of tasks oriented to measuring the same construct, based on the epistemological traditions of different cultural groups. Some questions may be more appropriate for the indigenous cultures in each nation, and could be used for indigenous as well as non-indigenous students. These questions would be accompanied by others that measure the same construct with an occidental focus of formal education. Thus the performance of students from different cultures could be compared on the same construct—a culturally adapted construct that could throw light on the magnitude of the bias of standardized examinations, as well as ways of adapting them to contexts of cultural diversity.

Current standardized examinations could be improved if their compatibility in indigenous contexts were analyzed. An idea that could give immediate results is the inclusion of a reduced number of open-ended questions on an exam given to a sample of indigenous students, along with another sample of non-indigenous students. The answers to open-ended questions would permit identifying cognitive processes and clarifying the meaning of constructs in both groups. Questions similar to multiple-choice questions can also be used for both samples. Comparing the answers to open-ended questions and multiple-choice questions in the two groups of students would permit greater understanding of the cognitive processes that each question triggers in both indigenous and non-indigenous students. Information would be provided about the differences in answers that can be attributed to cultural factors. In addition, the exercise would disclose the effect of the exam's format on the results of both populations. This simple idea could help to improve examinations as well as inferences based on the measurement of student achievement.

Further research is required in the Latin American context to explore the topic of bias in examinations. In countries like the United States, the analysis of bias is a common practice in the academic field as well as in educational policy (Freedle, 2003). Throughout this research, no studies of bias on standardized examinations were found in the Latin American context. This absence is indicative of the topic's limited inclusion in the agenda of the region's research.

Conclusion

The practice of educational evaluation in Latin America has acquired growing importance in the past thirteen years. The use of standardized examinations for making decisions is also a progressive phenomenon. Educational evaluation has flourished in a context of the legal recognition of cultural diversity in countries, causing tension between the measurement of "standard" knowledge and the acceptance of the diversity of knowledge in educational discourse. This tension has yet to be resolved.

The combination of factors like social inequalities, encouragement for high-stakes evaluation, and a lack of knowledge of the meaning of indigenous students' results on standardized tests could further limit the opportunities of indigenous students, the most disadvantaged group in the most unequal region in the world.

Standardized evaluation in indigenous contexts is problematic and leads necessarily to the question of appropriateness. Since the contrast between the epistemology of examinations and the epistemology of many indigenous cultures is apparently inevitable, the evaluation of indigenous students through standardized examinations might seem futile. Indigenous cultures promote types of knowledge, styles of teaching and learning, and methods of evaluation that often contradict the practices of formal education. Therefore, the epistemological clash could invalidate the evaluations' results.

In view of the challenges faced by the measurement of educational achievement in indigenous communities, this topic would seem to fit perfectly into the field of cross-cultural psychology. The problem of categories, types of intelligence, and the assumed universality of constructs are three topics of cross-cultural psychology that are closely linked to the measurement of educational performance in indigenous contexts. The scientific field of educational measurement would be benefited enormously by learning from the early errors of cross-cultural psychology.

The results of indigenous students on standardized examinations should be considered with extreme precaution. The differences between indigenous language and culture and those used on the exam, could impose limitations on the inferences based on exam results, and results should not be used to make individual judgments. Such judgments are not guaranteed, due to the technical characteristics of the tests. Examinations could be used to identify areas of opportunity for improving teaching; in addition, if they are accompanied by surveys of associated factors, joint analysis could permit improved understanding of the relation between student performance and explanatory variables of the population and school. Exam results should not be used to attribute causality, since they describe only a situation.

Given the structures of the educational systems of Latin America, the results of standardized examinations could be a variable that anticipates the future educational opportunities of indigenous students. Since bilingual or multicultural education does not extend beyond elementary school in most educational systems, indigenous students must handle the curriculum and dominant language in order to advance to higher levels in the educational system. In this context, tests that use the language of the dominant culture help to identify the gaps of inequality that affect indigenous children.

Educational evaluation in intercultural contexts is still weak in the region, since areas of evaluation in many nations are still in the process of consolidation. In spite of the presence of the technical ability for intercultural evaluation, political and financial support is weak.

Although standardized evaluation may help to document the educational inequalities between indigenous and non-indigenous students, questions regarding the degree of validity of

this type of evaluation in contexts of diversity remain unanswered. Questioning the validity of standardized examinations goes hand in hand with the use and possible consequences of exam results on the lives of indigenous students. In spite of the existence of media for graduating the validity of inferences based on test results, no solution is provided for the underlying problem of using examinations as a mechanism for unfairly limiting the educational opportunities of indigenous students, as well as the clash of perspectives on learning and knowledge between indigenous cultures and the culture represented on the examination.

In summary, this document has attempted to establish the bases for discussion on standardized evaluation in contexts of cultural diversity, and its relation with educational inequalities. At present, there are more questions than answers in this field of research, given the lack of basic data on the levels of bias in examinations. In Latin America, a long road remains to be traveled in researching the measurement of performance, cultural diversity, and equity. More empirical research will be necessary to guide our understanding of the topic in the region, and to shed light on the most adequate alternatives for action in the field of evaluation.

Notes

[1] I extend my deepest appreciation to Guillermo Solano-Flores, Luz María Moreno, Paula Tapia and two anonymous readers, whose comments on previous versions of this document were very valuable for its improvement. Any imprecision or error in this study is totally my responsibility.

[2] Chile is an exception in Latin America in at least three aspects. In first place, the System for Measuring the Quality of Education—SIMCE—publishes annual information on the performance of elementary and secondary schools, with the intent of generating public pressure on them to improve. In second place, Chile's Ministry of Education uses the data produced by SIMCE to distribute in each region of the country, additional resources to schools with low performance, in a desire for compensatory consequences. Lastly, the data from SIMCE, along with other information, have been used to award merit prizes to school teachers (McMeekin, 2000).

[3] This document uses the terms, occidental, Eurocentric, and dominant culture, as synonyms.

Bibliography

AERA, APA and NCME (1999). *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.

Apple, M. W. (2001). "Comparing Neo-liberal Projects and Inequality in Education", *Comparative Education*, vol. 37, number 4, pp. 409-423.

Barrera, S. (1998). *Sistema de medición y evaluación de la calidad de la educación (SIMECAL) y la modalidad intercultural bilingüe en Bolivia (EIB)*, unpublished, Bolivia.

Benadusi, L. (2001). "Equity and Education. A Critical Review of Sociological Research and Thought", in Hutmacher *et al.* (eds.) *In pursuit of equity in education,* Dordrecth/Boston/ London: Kluwer Academic Publishers.

Bransford, J.; Brown, A. and Cocking, R. (eds.) (2000). *How People Learn: Brain, Mind, Experience, and School.* Expanded edition. Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC.

Camilli, G. and Shepard, L. (1994). "Methods for Identifying Biased Test Items", *Measurement Methods for the Social Sciences Series*, vol. 4, Thousand Oaks, CA: Sage Publications.

Cheng, K.M. (1996). "Quality Assurance in Education: The East Asian Perspective", in K. Watson, S. Modgil and C. Modgil (eds.) *Educational Dilemmas: Debate and Diversity,* Volume 4: Quality in Education, London: Cassell.

Cizek, G. (1998). *Putting Standardized Tests to Test*, Washington, DC: Thomas B. Fordham Foundation.

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*, Cambridge, MA: Harvard University Press.

De Haan, M. (1998-1999). *Learning as Cultural Practice: How Children Learn in a Mexican Mazahua Community*, Amsterdam: Thela Thesis Publishers.

Diamond, J. and Spillane, J. (2004). "High-stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?", *Teachers College Record*, vol. 106, number 6, pp. 1145-1176.

DOE (2003). *No Child Left Behind*, < http://www.nclb.gov/> Quoted April 21, 2003.

Falcon, R. (2002). *México descalzo*, México: Plaza Janés.

Freedle, R. O. (2003). "Correcting the SAT's Ethnic and Social-class Bias: A Method for Re-estimating SAT Scores", *Harvard Educational Review*, vol. 73, number 1, Spring.

Fusarelli, L. (2004). "The Potential Impact of the No Child Left Behind Act on Equity and Diversity in American Education", *Educational Policy*, vol. 18, number 1, January-March.

Gopaul-McNicol, S. and Armour-Thomas, E. (2002). *Assessment and Culture: Psychological Tests with Minority Populations*, San Diego, CA: Academic Press.

Grisay, A. (2003). "Translation Procedures in OECD/PISA 2000 International Assessment", *Language and Testing,* vol. 20, number. 2, pp. 225-240.

Gronmo, L.; Kjaernsli, M. and Lie, S. (2004). "Looking for Cultural and Geographical Factors in Patterns of Responses to TIMSS Items", in Papanastasiou (ed.) *Proceedings of the IRC-2004 TIMSS,* Cyprus: Cyprus University Press, pp. 99-112.

Hall, G. and Patrinos, H. (at press). *Indigenous Peoples, Poverty and Human Development in Latin America 1994-2004*, Washington, DC: World Bank.

Hambleton, R. and Patsula, L. (2000). *Adapting Tests for Use in Multiple Languages and Cultures,* Laboratory of Psychometric and Evaluation Research Report, Amherst, Massachusetts: Massachusetts University.

Hambleton, R. (1993). *Translating Achievement Tests for Use in Cross-national Studies*, Amherst, Massachusetts: International Association for the Evaluation of Educational Achievement and National Center for Educational Statistics.

Hamilton, L. and Koretz, D. (2002). "Chapter 2: About Tests and Their Use in Test-based Accountability Systems", in Hamilton, *et al.* (eds.) *Making Sense of Test-based Accountability in Education*, Santa Monica, CA: Rand.

Helms-Lorenz, M.; Van de Vijver, F. and Poortinga, Y. (2003). "Cross-cultural Differences in Cognitive Performance and Spearman's Hypothesis: *G* or *C*?", *Intelligence*, vol. 31, pp. 9-29.

IADB (2004). *Profile: Strategic Framework for Indigenous Development.* Indigenous Peoples and Community Development Unit. Inter-American Development Bank. <http://www.iadb.org/sds/doc/IND%2DGN2295E2.pdf> Quoted September 15, 2004.

ILO (2003). *Indigenous and Tribal Peoples Convention* <http://www.ilo.org/ilolex/english/convdisp1.htm> /Quoted May 6, 2003.

INEGI (2000). *XII Censo General de Población y Vivienda*, México: INEGI.

Joe, J. (1994). "Revaluing Native-American Concepts of Development and Education", in Greenfield and Cocking (eds.) *Cross-cultural Roots of Minority Child Development.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Kalantzis, M.; Cope, B. and Slade, D. (1989). *Minority Languages and Dominant Culture: Issues of Education, Assessment and Social Equity,* London/New York. The Falmer Press.

Kirby, S. and Stecher, B (2004). "Chapter One: Introduction", in Stecher and Kirby (eds.) *Organizational Improvement and Accountability: Lessons for Education from Other Sectors*, Santa Monica, CA: Rand.

Kudo, I. (2002). *Indigenous Education in Peru: When Opportunity Speaks One Language.* Santiago, Chile: UNESCO-OREALC.

LLECE (1998). *Primer Informe de resultados.* Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grados de la educación básica, Santiago, Chile: UNESCO-OREALC.

Maxwell, B. (1996). "Translation and Cultural Adaptation of the Survey Instruments", in Martin and Kelly (eds.) *Third International Mathematics and Science Study (TIMSS) Technical Report,* volume I: *Design and Development*, Chestnut Hill, MA: Boston College.

McMeekin, R. (2000). "Implementing School-based Merit Awards: Chile's Experience. Technical Notes", *Education Reform and Management Publication Series*, vol. III, number 1, June.

McNeil, L. (2000). "Creating New Inequalities: Contradictions of Reform", *Phi Delta Kappan,* pp. 729-734, June.

McQueen, J. and Mendelovits, J. (2003). "PISA Reading: Cultural Equivalence in a Cross-cultural Study", *Language and Testing*, vol. 20, number 2, pp. 208-224.

Messick, S. (1993). "Validity", in Linn (ed.) *Educational Measurement*, New York: National Council on Measurement in Education and American Council on Education.

Mier y Terán, M. and Romero, C. (2003). "Inequalities in Mexican Children's Schooling", *Journal of Comparative Family Studies,* 34(3): 435-454.

Muller, C. and Schiller, K. (2000). "Leveling the Playing Field? Students' Educational Attainment and States' Performance Testing", *Sociology of Education*, vol. 73, pp. 196-218.

NCR-National Research Council (1997). *Educating Language-minority Children*, Washington, DC: National Academy Press.

Padilla, A. (2001). "Issues in Culturally Appropriate Assessment", in L. Suzuki, J. Ponterorro and P. Meller (eds.) *Handbook of Multicultural Assessment: Clinical, Psychological and Educational Applications,* San Francisco, CA: Jossey-Bass.

Popham, W.J. (2000). *Modern Educational Measurement: Practical Guidelines for Educational Leaders*, 3rd ed., Boston: Allyn & Bacon.

Price-Williams, D. (1975). *Explorations in Cross-cultural Psychology*, San Francisco, CA: Chandler & Sharp Publishers.

Ravela, P. (2001). *¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?* Santiago, Chile: PREAL.

Roderick, M. *et al.* (1999). *Ending Social Promotion: Results from the First Two Years*, Chicago, Illinois: Consortium on Chicago School Research.

Sacks, P. (1999). *Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It,* Cambridge, Mass.: Perseus Books.

Solano-Flores, G. and Nelson-Barber, S. (2001). "On the Cultural Validity of Science Assessments", *Journal of Research in Science Teaching*, vol. 38, number 5, pp. 553-573

Suárez-Orozco, M. (1994). "Remaking of Psychological Anthropology", in Suárez-Orozco, *et al. The Making of Psychological Anthropology II.* Forth Worth, Texas: Hartcourt Brace College Publishers.

Tharpe, R. (1994). "Intergroup Differences among Native Americans in Socialization and Child Cognition: An Ethnogenetic Analysis", in Greenfield and Cocking (eds.) *Cross-cultural Roots of Minority Child Development*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Tirado, F., Backhoff, E., Larrazolo, N. y Rosas, M. (1997). "Validez predictiva del examen de habilidades y conocimientos básicos (EXHCOBA)", *Revista Mexicana de Investigación Educativa,* enero-junio 1997, vol 2, núm. 3, pp. 67-84.

Treviño, E. y G. Treviño (2004). *Estudio sobre las desigualdades educativas en México: la incidencia de la escuela en el desempeño académico de los alumnos y el rol de los docentes*, colección Cuadernos de investigación, núm 5, México: Instituto Nacional para la Evaluación de la Educación.

Treviño, E. (en prensa*). Buenas prácticas en el uso de evaluaciones educativas para decisiones de política en América Latina*, Santiago, Chile: UNESCO-OREALC.

UNESCO (2000). *Informe subregional de América Latina: Evaluación de educación para todos en el año 2000*, Santiago, Chile: OREALC.

Van de Vijver, F. (1997). "Meta-analysis of Cross-cultural Comparisons of Cognitive Test {erformance", *Journal of Cross-Cultural Psychology,* vol. 28, number 6, November, pp. 678-709.

Werner, E. (1979). *Cross-cultural Child Development: A View from the Planet Earth,* Monterey, CA: Brooks/Cole Publishing Co.

Winkler, D. (2004). "Investigaciones sobre etnicidad, raza, género y educación en las Américas", en Winkler y Cueto (eds.) *Etnicidad, raza, género y educación en América Latina,* Santiago, Chile: PREAL.

Wolff, L. (1998). *Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos*, Washington, DC/Santiago, Chile: PREAL.