Conversations on Validity Around the World

Comments on Validity Issues in International Large Scale Assessments

Eduardo Backhoff, México
Introduction

In his presentation, Dr. Michael Feuer established, among others, the following premises:

1. Recently, the number of countries undertaking large scale student assessments raised up to 67.
2. International student surveys, as TIMSS and PISA, have made countries put greater attention on the quality of their education systems.
3. Student assessment programs are becoming more known by politicians, by society in general and, especially, by the mass media (MM).
4. MM have published sensationalist news to call the public attention and, in turn, it has pressed the government to improve what international assessments point to as weaknesses of educational systems.
Feuer’s approach

- The real value of these programs is not well documented by empirical information available.
- There is no data supporting the direct relationship between the PISA results and the economic growth of countries.
- The central validity of this reasoning rests on two conceptual pillars:
  1. The substantive theory linking the effects of education on the economy of countries.
  2. The proper interpretation of empirical information available on education and economy.
On the information available from the U.S. economy, Feuer stated:

- In recent decades, productivity in manufacturing has been stable and slightly upward.
- The total productivity growth is very strong.
- Accordingly, the U.S. economy is not in a serious condition.
- Considering reliable assessment of students, the U.S. is not declining.
- In TIMSS, the performance in math has been improving; NAEP, confirms this story; in PISA, the % students with higher performances is better in the U.S. than in 5 of 14 countries with the highest rankings.
The role of Mexico in international and national assessment surveys

- Has participated in PEARLS and in several studies of UNESCO.
- In 2002, it was created the National Institute for Educational Evaluation (INEE), a NAEP-like institution, to assess students from preschool through high school.
- In 2005, the Ministry of Education created the program called ENLACE: National Assessment for Educational Centers, which yearly assess, in a census way, almost all basic education students from different grades.
- Additionally, many states are beginning to create their own centers of student’s assessment.
Three major impacts of international assessments in Mexico

1. An increased participation in student assessment programs, primarily using over sampling procedures.
2. To established country government goals to increase nation PISA scores, in an unrealistic periods of time.
3. To establish educational reforms, both in curriculum and teachers professional development, as well as to implement a program called Universal Teacher Assessment of Basic Education (a bonus payment program).
Mass media and private sector pressuring on Mexico’s education system

- Mass media have put too much attention on PISA results, occupying the front pages of the most prestigious newspapers in the country.
- Big private organizations have been pressing the authorities to improve the quality of educational services.
- One example is the recent Mexican documentary film "De panzazo", which describes the educational problems of the country, supported mainly by PISA test results, and concluding that the only way to improve education in Mexico is through testing teachers, with severe consequences: to rise salaries to those with the highest results and dismiss teachers who obtain the lowest.
The interpretation validity of international student assessment: the case of Mexico (1)

- Some studies on the translation and adaptation of international tests (TIMSS and PISA), have demonstrated that a large amount of their items suffer from significant problems that limit their interpretive validity.
- Some studies have demonstrated the limited cultural validity of context questionnaires, which serve to interpret the results of international student assessment studies.
The interpretation validity of international student assessment: the case of Mexico (2)

- The lack of statistical invariance in TALIS countries means that comparisons cannot be made between them in some school variables. This was the case of teacher beliefs, teaching practices and teacher professional collaboration.

- In some countries, unexpected changes in PISA scores from one year to another have been the subject of various interpretations.

- These changes are observed more in countries with low scores, for which the international tests have not enough items.

- One case was Brazil, in 2006, that decreased its scores in math but not in reading. Another case was Mexico in 2009 that showed a significant increase in reading but not math.

- One hypothesis for both cases is that these changes are the result of test construction problems, rather than real changes in education systems (as politicians want to believe).
Conclusions

1. As Feuer stated, international comparative assessments must have better theoretical framework of validity, as well as empirical evidence supporting decision-making, based on:
   - International rankings.
   - Comparative scores among countries.

2. In my own experience we must carefully interpret:
   - Countries’ score changes over time.
   - Countries’ context questionnaires results.
   - The equivalence of translated/adapted test and questionnaires items.
   - The appropriateness of international organization recommendations for educational changes. Some school variables work differently in each country, depending on its own social and educational systems.
Frequency of PISA-2006 items (n=193) with translation errors by error dimension

Source: Backhoff, Contreras-Niño & Solano-Flores. (2011) La Teoría del Error de Traducción de Pruebas y las evaluaciones internacionales de TIMSS y PISA. México: INEE.
Percentage of PISA-2006 items (n=193) with at least one error on each dimension

<table>
<thead>
<tr>
<th>Translation error dimension</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Style</td>
<td>48</td>
</tr>
<tr>
<td>Format</td>
<td>53</td>
</tr>
<tr>
<td>Conventions</td>
<td>3</td>
</tr>
<tr>
<td>Information</td>
<td>53</td>
</tr>
<tr>
<td>Grammar</td>
<td>53</td>
</tr>
<tr>
<td>Semantics</td>
<td>78</td>
</tr>
<tr>
<td>Construct</td>
<td>35</td>
</tr>
<tr>
<td>Register</td>
<td>21</td>
</tr>
<tr>
<td>Culture</td>
<td>5</td>
</tr>
<tr>
<td>Origin</td>
<td>41</td>
</tr>
</tbody>
</table>

Correlation between errors in language and content dimensions and item p-value
\[ r = -0.404 \]

## Summary of Predicting and Control Variables correlations with PISA results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Korea</th>
<th>Mexico</th>
<th>U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicting variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEADINS</td>
<td>Instructional leadership</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LEADTCH</td>
<td>Leadership of teachers</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>LEADPD</td>
<td>Leadership for professional development</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>AUTOFNCE,</td>
<td>School financial autonomy</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>AUTOTCH</td>
<td>School autonomy in teaching and learning</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TSTSCHACC,</td>
<td>Testing for accountability</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>TSTCOMP</td>
<td>Testing to compare</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>TSTINS</td>
<td>Testing to improve instruction</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Control variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPTYHIND</td>
<td>School instruction hindered by shortage of human and material resources</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>COMMTY</td>
<td>Community size where school is located</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>OUTSCH</td>
<td>Student attend enrichment lessons in test language</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ESCS</td>
<td>Student socio-economic status</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Source: Seon, Slater, and Backhoff (in press). Principal Perceptions and Student Achievement in Reading in Korea, Mexico and the United States: Educational Administration Quarterly.
TALIS México report

- Family and student characteristics were stronger associated with learning results than school characteristics.
- Regarding teachers beliefs, constructivist vision is positively related with learning results, while direct teaching is negatively associated.
- However, student-oriented pedagogic practices (associated with a constructivist vision) showed negative relationship with ENLACE results.

Source: Backhoff, Andrade, Bouzas, Santos del Real & Santibañez (2009). Estudio Internacional sobre la Enseñanza y el Aprendizaje (TALIS): resultados de México