DETAILS OF A NON-RANDOMIZED EVALUATION STUDY
http://www.romadecade.org/a_successful_school_integration_program_2009

European Commission, DG Regional Policy
Impact Evaluation Spring School

Gabor Kezdi
CEU

Lake Balaton, May 2011
The program

- National Integrated Education Network (Országos Oktatási Integrációs Hálózat, OOIH)
  - organized by the Hungarian Government
  - the evaluated program started in fall 2003
  - in 45 primary schools a.k.a. "base schools"
  - grandfathered, starting with those in grades 1 and 5 in 2003/4

- Goal: integrated and quality education of Romani and/or disadvantaged students
  - participating schools have sizeable Romani/disadvantaged minority
  - integration required across and within groups
  - financial support: equipment, physical environment, etc.
  - educational support: intensive training in modern methods, consulting, etc.
  - multiple elements, freedom of choice for schools
  - emphasis on schools (as opposed to individual teachers)
  - monitoring (by young Roma)
Evaluation questions

1. Effects on schools
   - integrated education
   - education methods, classroom management
   - student behavior

2. Effects on students
   - reading tests, further for further education after grade 8
     * the usual outcomes examined in evaluations of this kind
   - self esteem, locus of control, coping
     * more specific to this evaluation
     * recent labor economics research show they are very important for success of disadvantaged people
     * development research shows programs of this kind can affect them
     * interviews with program administrators & teachers showed these were important targets
   - social distance across ethnic groups (Roma versus non-Roma)
     * obvious outcome of school integration
Non-experimental design

- Short of randomization ideal method would have been difference-in-differences on matched sample
  - i.e. compare how students develop in program schools vs. matched control schools
- Needed control group for many reasons
  - potential secular trends in student outcomes
  - potential before-program effects
- Needed to assess development of students
  - simply comparing results after the program may be misleading
  - in case matching is imperfect, differences in students may cause problems
  - comparison of results to before-program results can help a lot
- Matched control group at what level?
  - schools or students?
- How could we get before-program measures?
- And what kinds of measures?
Control group selection

Matching at school level

- 60 schools: 30 program schools and 30 control schools
- matched sample: every program school has a matched control pair
  - within same region and village/town type
  - propensity score matching using many pre-program characteristics (student composition, competence test results, some evidence on Romani students etc.)
  - checked with sociologists with local knowledge

Student sample \((n \approx 4000)\)

- all students who were in grade 1 or 5 in 2003/4
- followed them through grade 4 and 8, respectively
  - following students in an anonymous way was tricky
  - kept track of their entry number of class register and harmonized registers across schoolyears
## Variables used for the matching

Measured in 1999 and 2003

<table>
<thead>
<tr>
<th>Variable</th>
<th>Program schools</th>
<th>Control schools</th>
<th>National average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population in town/village of school</td>
<td>87,363</td>
<td>83,399</td>
<td>225,992</td>
</tr>
<tr>
<td>School size (number of students)</td>
<td>320</td>
<td>311</td>
<td>286</td>
</tr>
<tr>
<td>Fraction eligible for Roma minority support in 1999 (%)</td>
<td>31</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td>Fraction of students at risk (%)</td>
<td>16</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Fraction of students (%) with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother’s education less than 8 grades</td>
<td>9</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Mother’s education exactly 8 grades</td>
<td>35</td>
<td>33</td>
<td>21</td>
</tr>
<tr>
<td>Father’s education less than 8 grades</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Father’s education exactly 8 grades</td>
<td>26</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>No working parent</td>
<td>29</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Number of books at home 0 to 50</td>
<td>29</td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>Number of books at home approximately 50</td>
<td>16</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Competence scores at school level, 6th grade, spring 2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics: school average</td>
<td>456</td>
<td>446</td>
<td>500</td>
</tr>
<tr>
<td>Reading: school average</td>
<td>449</td>
<td>436</td>
<td>500</td>
</tr>
<tr>
<td>Mathematics: school standard deviation</td>
<td>88</td>
<td>88</td>
<td>87</td>
</tr>
<tr>
<td>Reading: school standard deviation</td>
<td>92</td>
<td>95</td>
<td>89</td>
</tr>
</tbody>
</table>
Geographic distribution of the matched sample
## Characteristics of the matched sample

**Student composition, measured in 2006**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Grade 3</th>
<th></th>
<th></th>
<th></th>
<th>Grade 7</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Program</td>
<td>Control</td>
<td>National</td>
<td>Program</td>
<td>Control</td>
<td>National</td>
<td></td>
</tr>
<tr>
<td>Number of students with information</td>
<td>810</td>
<td>909</td>
<td>91,349</td>
<td>757</td>
<td>897</td>
<td>92,588</td>
<td></td>
</tr>
<tr>
<td><strong>Fraction girls (%)</strong></td>
<td>47</td>
<td>48</td>
<td>50</td>
<td>49</td>
<td>50</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Welfare eligible (%)</td>
<td>37</td>
<td>37</td>
<td>19</td>
<td>31</td>
<td>35</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Eligible for free meal in school (%)</td>
<td>59</td>
<td>60</td>
<td>28</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Eligible for free school books (%)</td>
<td>73</td>
<td>83</td>
<td>57</td>
<td>66</td>
<td>78</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Average family size</td>
<td>4.9</td>
<td>5.0</td>
<td>4.5</td>
<td>4.7</td>
<td>4.8</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td><strong>Fraction with father in family (%)</strong></td>
<td>78</td>
<td>76</td>
<td>83</td>
<td>76</td>
<td>73</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Mother uneducated (8 grades or less, %)</td>
<td>43</td>
<td>43</td>
<td>21</td>
<td>40</td>
<td>40</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Father uneducated (8 grades or less, %)</td>
<td>35</td>
<td>35</td>
<td>16</td>
<td>32</td>
<td>29</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Fraction with working mother (%)</td>
<td>41</td>
<td>42</td>
<td>66</td>
<td>52</td>
<td>51</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Fraction with working father (%)</td>
<td>64</td>
<td>64</td>
<td>84</td>
<td>69</td>
<td>65</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Avg. size of apartment (sq meter/capita)</td>
<td>19.0</td>
<td>19.0</td>
<td>n.a.</td>
<td>20.8</td>
<td>20.5</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td><strong>Average monthly spending (HUF '000)</strong></td>
<td>109</td>
<td>110</td>
<td>n.a.</td>
<td>120</td>
<td>115</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>Fraction who think poor (%)</td>
<td>38</td>
<td>38</td>
<td>17</td>
<td>29</td>
<td>29</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Fraction who were not on holiday (%)</td>
<td>41</td>
<td>41</td>
<td>21</td>
<td>28</td>
<td>31</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Fraction without a car (%)</td>
<td>46</td>
<td>48</td>
<td>27</td>
<td>44</td>
<td>43</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>0 to 50 books at home (%)</td>
<td>35</td>
<td>35</td>
<td>16</td>
<td>29</td>
<td>27</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>
Data collections

Three "waves"

1 2004/5: "input measurement"
   - students: cognitive and non-cognitive skills, grades, etc.
     - self-administered questionnaires to students, with interviewer help
   - schools: class composition, participation in other programs etc.
     - interviews with schoolteachers and directors

2 2005/6:
   - students: family background
     - self-administered questionnaire to be filled out at home with parents
   - classroom observations
     - standardized questionnaires filled out by trained observers

3 2006/7: "outcome measurement"
   - cognitive and non-cognitive skills etc.
     - self-administered questionnaires to students, with interviewer help
   - students’ ethnic identity (see later)
   - direction of further education after grade 8
     - admission results collected from schools after end of schoolyear
Methods of data analysis

- Simple comparisons of "after-program" (wave 3) results
  - difference of average results in program schools versus average results in control schools
  - overall difference, differences by ethnicity and disadvantaged status

- Regressions controlling for earlier results and many other things
  - family background
  - results of outcomes measured two years back
    - short of true "before-program" measures this is the closest we could do
    - in principle this would "overdo" the controlling
  - social desirability (desire to make good impression)

- Split sample results
  - restricted to program schools that were not integrating before the program, and comparing their results to their own control schools
  - reason: less confounded by integrated education before the program
Measurement of ethnic identity

- Assessed by parents
  
  "Please indicate which of the following characterizes the ethnic/national identity of your child. Feel free to indicate multiple categories. My child is a) Hungarian, non-Roma; b) Hungarian, Roma; c) Hungarian, partly Roma; d) Non-Hungarian, Roma; e) Romanian; f) Slovak;..."

---

Fraction of Romani students from parents’ declaration compared to schoolteachers’ estimates

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>C</td>
</tr>
<tr>
<td>Fraction Roma, parental declaration (%)</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Missing parental declaration (%)</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Fraction Roma, schoolteacher estimates (%)</td>
<td>34</td>
<td>36</td>
</tr>
</tbody>
</table>
Results presented here

- Simple comparisons of average student outcomes in program schools versus control schools
  - "after-program" (wave 3) measures

- Regression estimates of the treatment effect after controlling for many things
  - i.e. program school versus control school differences after controlling for many things
  - those many things include results two years earlier
  - generalized difference-in-differences interpretation

- Split sample results
  - regression results from program schools that were not integrating before the program, and comparing their results to their own control schools

- One set of outcomes will be presented
"Non-cognitive skills"

- **Self-esteem**
  - measurement: adopted and standardized Harter-SPPC test
  - 4+1 components (exterior, school performance, good behavior, social functioning) + general,
  - each component is a two-item test
  - we sometime aggregate all into one score

- **Locus of control**
  - the extent to which people think they can influence their own destiny
  - measurement: adopted and standardized short Rotter test (four items)

- **Coping**
  - the extent to which people can go on w/o lasting negative consequences after some difficult situations even if the problem remains
  - measurement: standardized test of our own

- All are measured by standardized (0,1) test scores
Simple comparisons by ethnicity

- Locus of control, Roma
- Locus of control, non-Roma
- Self-esteem, Roma
- Self-esteem, non-Roma
- Coping, Roma
- Coping, non-Roma

Program schools vs Control schools
Regression results (also from restricted sample)
Locus of control. Coefficient estimates of treatment effect.

Table 6.2 - Program versus control differences in internal locus of control. Regression results, controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure from a four-item Rotter-scale.

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Roma students</th>
<th>Non-roma students</th>
<th>Disadvantaged students</th>
<th>Non-disadvantaged students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>+0.13**</td>
<td>+0.24**</td>
<td>+0.10</td>
<td>+0.15+</td>
<td>+0.11+</td>
</tr>
<tr>
<td>Narrow sample</td>
<td>+0.10+</td>
<td>+0.27+</td>
<td>+0.07</td>
<td>0.00</td>
<td>+0.10+</td>
</tr>
</tbody>
</table>

Standard errors are robust to heteroskedasticity:
+ Statistically significant at the 10% level.
* Statistically significant at the 5% level.
** Statistically significant at the 1% level.
Regression results (also from restricted sample)
Overall self-esteem score. Coefficient estimates of treatment effect.

Table 6.6 – Program versus control differences in positive self-esteem. Regression results controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure of the SPPC total score.

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Roma students</th>
<th>Non-roma students</th>
<th>Disadvantaged students</th>
<th>Non-disadvantaged students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>+0.24**</td>
<td>+0.28**</td>
<td>+0.24**</td>
<td>+0.38**</td>
<td>+0.20**</td>
</tr>
<tr>
<td>Narrow sample</td>
<td>+0.33**</td>
<td>+0.29+</td>
<td>+0.35**</td>
<td>+0.25+</td>
<td>+0.32**</td>
</tr>
</tbody>
</table>

Standard errors are robust to heteroskedasticity.
+ Statistically significant at the 10% level.
* Statistically significant at the 5% level.
** Statistically significant at the 1% level.
Regression results (also from restricted sample)

Coping. Coefficient estimates of treatment effect.

Table 6.8 – Program versus control differences in coping. Regression results controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure from a four-item coping scale.

<table>
<thead>
<tr>
<th></th>
<th>ALL STUDENTS</th>
<th>ROMA STUDENTS</th>
<th>NON-ROMA STUDENTS</th>
<th>DISADVANTAGED STUDENTS</th>
<th>NON-DISADVANTAGED STUDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>+0.08</td>
<td>+0.31*</td>
<td>+0.03</td>
<td>+0.13</td>
<td>+0.17</td>
</tr>
<tr>
<td>Narrow sample</td>
<td>+0.16+</td>
<td>+0.41*</td>
<td>+0.08</td>
<td>+0.12</td>
<td>+0.13</td>
</tr>
</tbody>
</table>

Standard errors are robust to heteroskedasticity.
+ Statistically significant at the 10% level.
* Statistically significant at the 5% level.
** Statistically significant at the 1% level.
Thank you for your attention