

Student Nonresponse Weighting Adjustments in PISA

Keith Rust, Westat and University of Maryland
Christian Monseur, Université de Liège
Sheila Krawchuk, Westat

PISA Research Conference
Kiel, Germany
14 September, 2009

Westat®

Overview

- PISA sample design
- Student weights for analysis
- School and student nonresponse adjustments
- Student nonresponse adjustment in 2003 (method A)
- Student nonresponse adjustment in 2006 (method B)
- Comparison of methods A and B
- Effect of change in methods on trend estimates
- Conclusions

Westat®

2

PISA Sample Design

- In each country, two-stage sample of schools and students
- School sample
 - Stratified using geographic and school characteristics
 - Schools selected with probability proportional to size
 - Size measure a function of estimated enrollment of eligible students

Westat®

3

PISA Sample Design (continued)

- Student sample
 - Equal probability systematic sample of 35 students within each school
 - Goal is to select an approximately equal probability sample of students within each school stratum

Westat®

4

Student Weights for Analysis

- Each student is assigned a survey weight, consisting of six components
 - school base weight
 - school weight trimming factor
 - school nonresponse adjustment factor
 - student weight within school
 - student nonresponse adjustment factor
 - Student weight trimming factor

Student Weights for Analysis

- School base weight: reciprocal of the school selection probability
- Student weight within school: reciprocal of the student selection probability, within school
- School weight trimming factor: reduces unduly large school weights to reduce variance; large weights arise from poor enrollment estimates or incorrect stratum classifications
- Student weight trimming factor: reduces unduly large overall student weights to reduce variance

School and Student Nonresponse Adjustments

- Nonresponse adjustments adjust the weights of the respondents such that, after the adjustment, the weighted distributions of the respondents match those of the full sample prior to the adjustment.
- Nonresponse adjustment cells define the level at which the adjustments are made.
- The sum of the adjusted weights of the respondents within the cell equals the sum of unadjusted weights of the full sample.
- Nonresponse adjustments reduce bias if the variables that define the cells are moderately or highly correlated with both:
 - the survey characteristic (student achievement)
 - response status

School and Student Nonresponse Adjustments

- School Nonresponse
 - Varies greatly by country
 - Sometimes varies greatly by stratum within country
 - Sometimes varies considerably over time within a country
- Therefore school nonresponse adjustment procedures are very country specific, and can vary considerably within country across PISA cycles.

School and Student Nonresponse Adjustments

- Student Nonresponse
 - does not vary much by country (85-95%)
 - does not vary much over time
 - limited data on nonrespondents
- Therefore student nonresponse adjustment procedures can be similar across countries, and need not necessarily vary over time.
- Within-school variance is often around 90% of total population variance (within strata). Thus in general student nonresponse has the greater potential to bias the results.
- In PISA, achievement is highly correlated with grade
- Girls are much better at PISA reading; boys are better at PISA mathematics

Student Nonresponse Adjustments in 2003 – Method A

- Adjustments based on, in priority order
 - school
 - grade
- Grade dichotomized (higher/lower) within each *country*
- Nonresponse classes required a minimum of ten respondents to avoid large variances

Student Nonresponse Adjustments in 2003 – Method A

- Often within a school most students in a single grade
- Result: often grade was largely not used in creating classes
- Thus, nonresponse bias reduction was largely obtained by controlling on school
 - often little response rate variation across schools
 - often little difference in achievement across schools

Student Nonresponse Adjustments in 2006 – Method B

- Adjustments based on
 - school nonresponse adjustment class (a group of several similar schools)
 - grade
 - gender
- Grade dichotomized within each *explicit stratum*
- Nonresponse classes required a minimum of 15 respondents
- Compared with 2003:
 - reduced reliance on school membership to reduce bias
 - increased use of both grade and gender
 - slightly increased adjustment class size

Comparison of Methods A and B

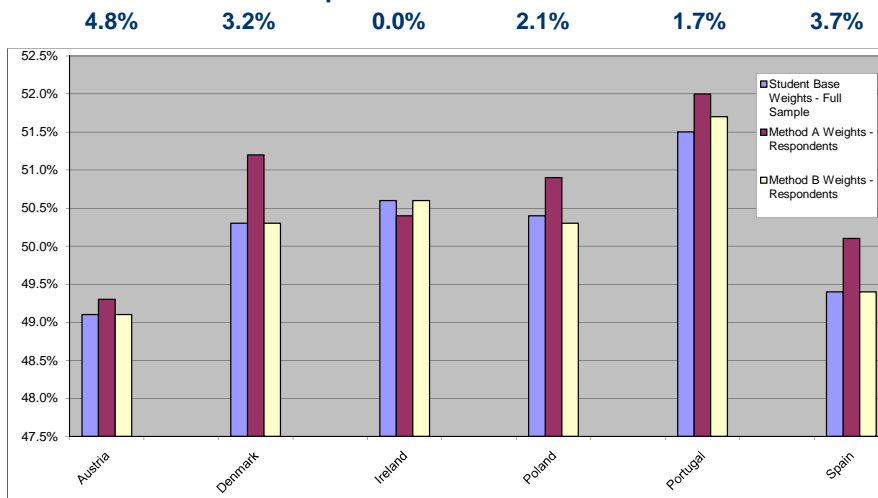
- Evaluated the impact of the change in procedures by repeating the adjustments for 2006 using method A and comparing the results of methods A and B
- Analysis done for all countries that also participated in 2003, as these might have trend estimates affected by the change in method

Comparison of Methods A and B

- Consider the differences in weight distribution by gender and grade, to determine whether method B was effective in reducing potential bias due to differential nonresponse by gender and grade
- Compare the point estimates and standard errors for 2006 achievement results using each method
- Evaluate the trend estimates from 2003 to 2006 for those countries where the two methods gave different results in 2006

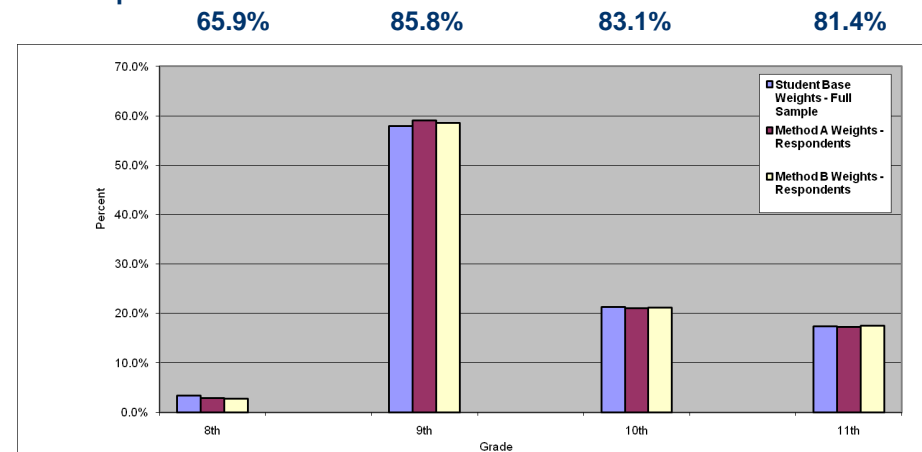
Percentage Girls, PISA 2006

Gender difference in response rate:



Response Rates and Weight Distribution by Grade PISA 2006, Ireland

Response Rate:



Response Rates and Weight Distribution by Grade PISA 2006, Italy

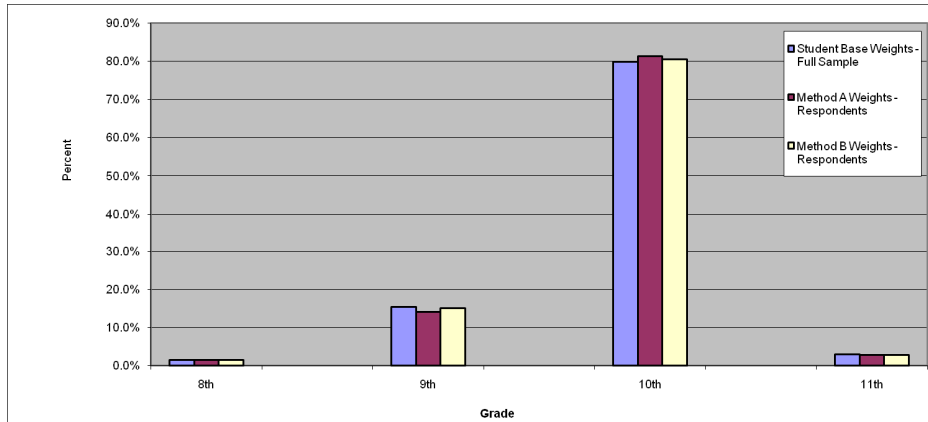
Response Rate:

82.8%

83.1%

94.2%

87.8%



Westat

17

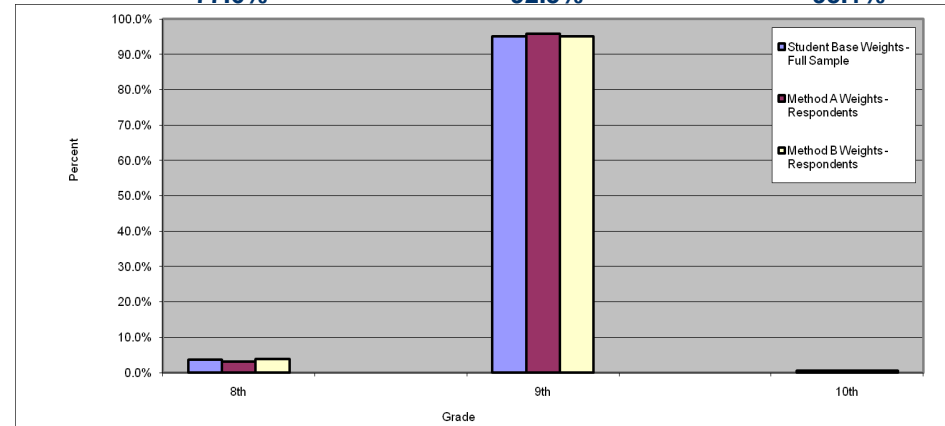
Response Rates and Weight Distribution by Grade PISA 2006, Poland

Response Rate:

77.6%

92.5%

95.1%



Westat

18

Response Rates and Weight Distribution by Grade PISA 2006, Portugal

Response Rate:

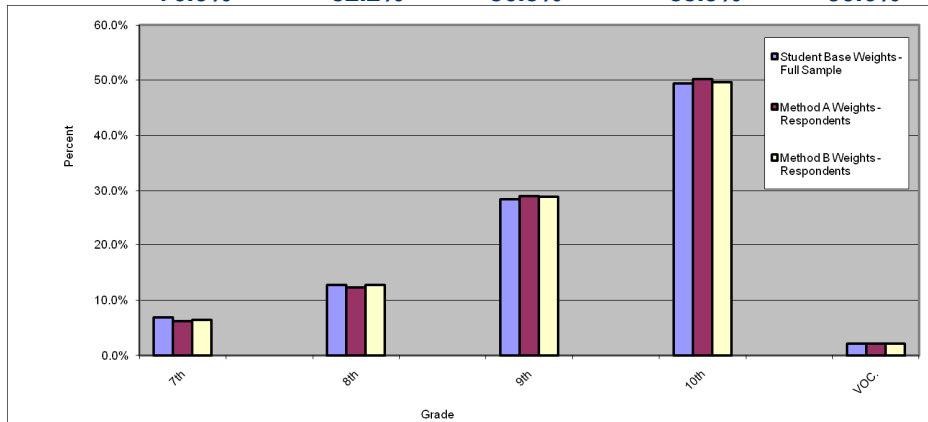
76.0%

82.2%

86.8%

88.9%

90.6%



Westat

19

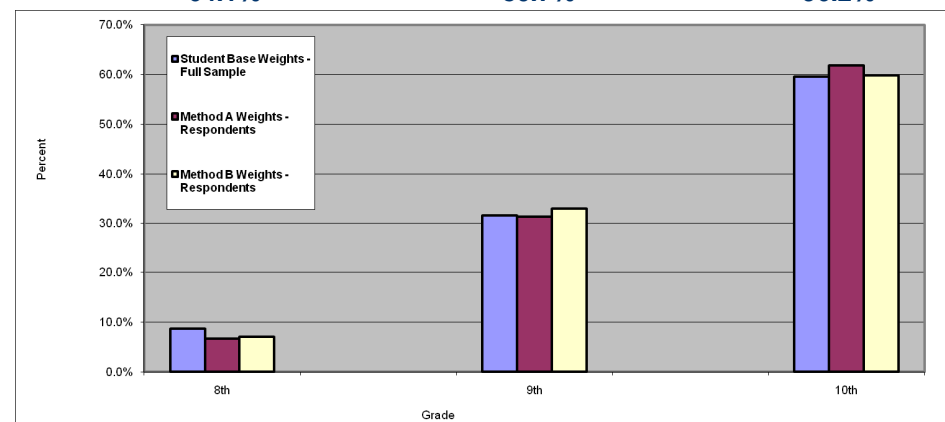
Response Rates and Weight Distribution by Grade PISA 2006, Spain

Response Rate:

64.1%

80.7%

96.2%



Westat

20

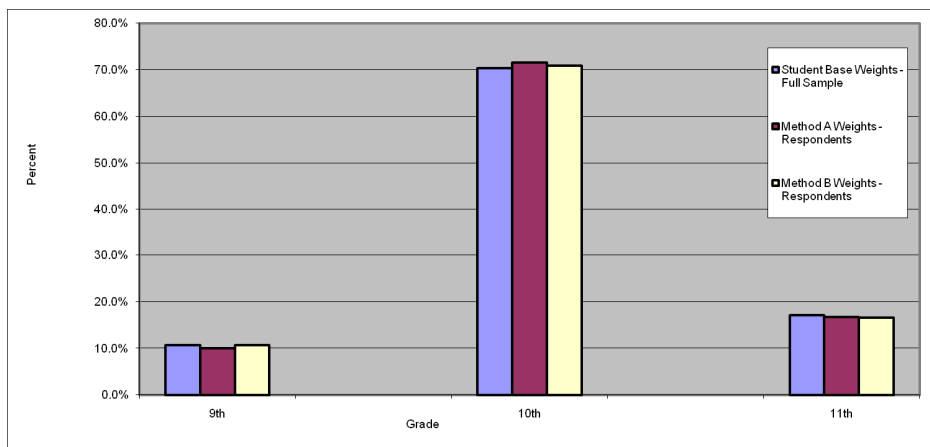
Response Rates and Weight Distribution by Grade PISA 2006, USA

Response Rate:

84.1%

92.5%

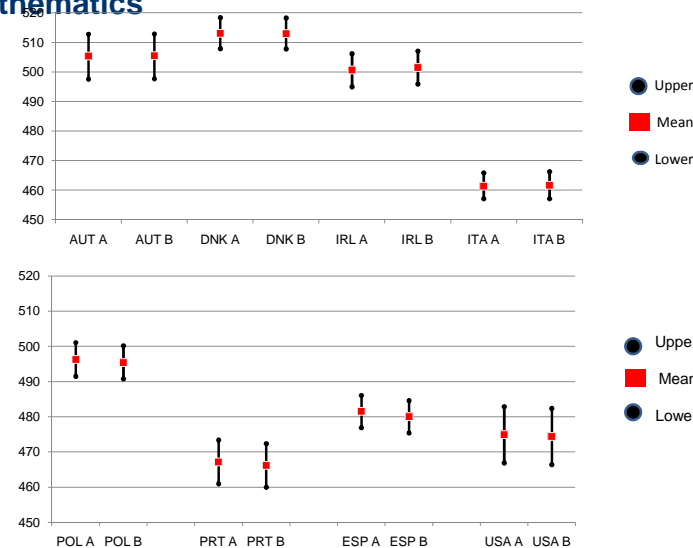
88.2%



Westat

21

Comparison of Methods A and B – PISA 2006 Mathematics



Westat

22

Comparison of Methods A and B - Summary

- Results for reading and science similar to mathematics
- Four countries differed by about one point or more
 - Ireland: +0.9 (math), +1.0 (reading), +1.0 (science)
 - Poland: -0.8, -1.4, -0.9
 - Portugal: -1.0, -1.2, -1.2
 - Spain: -1.6, -1.6, -1.5
- In each case, method B adjusted the gender and age distribution more effectively
- No difference in estimated standard errors
- Except for Spain, differences less than 0.5 standard errors
- Spain differences less than one standard error

Westat

23

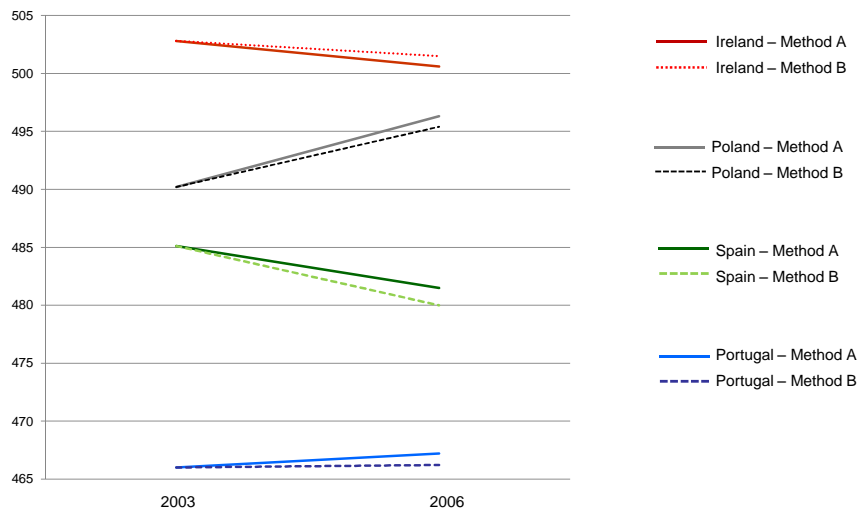
Comparison of Methods A and B - Summary

- For Ireland, method B decreased grade 9 (versus 10 and 11) and increased girls – increased achievement mean
- For Poland, method B decreased grade 9 (versus grade 8) and decreased girls – decreased achievement mean
- For Portugal, method B decreased grade 10 (versus 8 and 7) and decreased girls – decreased achievement mean
- For Spain, method B decreased grade 10 (versus 8 and 9) and decreased girls – decreased achievement mean
- For Spain, method B distribution between grades 8 and 9 suggests that positive bias remains
- Adjustments to grade distribution, although small, had the noticeable effects on achievement mean

Westat

24

Effect of Change in Methods on Trend Estimates – Mathematics



Effect of Change in Methods on Trend Estimates

- Change in method had effects that were:
 - large relative to the size of the estimated trend in Ireland and Portugal
 - small relative to the standard error of the trend - no significant change over time in Ireland, Spain, Portugal

Conclusions

- Method B, adopted in 2006, appears effective in reducing bias due to differential student nonresponse by grade and gender
- The change in method had a noticeable effect on the mean achievement results in a few countries, consistent with the differential response rates evident in those countries
- All changes were small, relative to sampling error, and very small in effect size terms (i.e. relative to the standard deviations of the achievement scales)
- There were no noticeable changes in sampling error estimates
- No artifactual trend results of any consequence resulted

Contact information

- Keith Rust
 - KeithRust@Westat.com
- Christian Monseur
 - CMonseur@ulg.ac.be
- Sheila Krawchuk
 - SheilaKrawchuk@Westat.com