A Response to Rosnick and Weisbrot

Daniel Ortega and Francisco Rodríguez
March 2008

In Rosnick and Weisbrot (2008) - henceforth RW - argue that our results in Ortega and Rodríguez (2008) – henceforth OR - where we found no consistent statistically significant effect of the Venezuelan literacy program on literacy outcomes, are sensitive to specification and rely on data that cannot adequately capture the effects of a large scale literacy program. In this short note we will show that RW's criticism is unfounded, and that the results of their analysis are inconsistent with the Venezuelan government's claim of illiteracy eradication.

In discussing RW's criticism, it is important to bear in mind that there are, broadly speaking, three possible hypotheses about the effects of Venezuela's *Robinson* literacy program which are amenable to testing using the Households survey. In one hypothesis, Venezuela had a massive literacy program in which upwards of one million previously illiterate persons were enrolled and saw significant improvements in their writing skills. This would be consistent with the Venezuelan government's initial claim of near-illiteracy eradication, which has been repeated in UNESCO (2006) and several academics, among them Weisbrot (2008, p. 22), who wrote that "over one million people participated in adult literacy programs". We will call this the **large program** hypothesis. A second hypothesis would be that Venezuela saw a moderately sized literacy program, in which tens of thousands of people were enrolled. We will call this the **small program hypothesis**. According to a third hypothesis, the Venezuelan government did not enact a literacy program of significant proportions, and any announcements that such a program existed were a farce. We will call this the **no program hypothesis**.

In Ortega and Rodríguez (2008), we present a battery of tests of the *Robinson* program, using time-series, cohort, and state-level data derived from the Households Survey. A large number of our coefficients are statistically insignificant – though some of them are significant and positive, such as our lagged coefficients on the oldest age cohorts reported in our Table 4. Based on that analysis, we conclude that "we find at most a small positive effect of *Robinson* on literacy rates, and in many specifications the program impact is statistically indistinguishable from zero…The results appear to be inconsistent with recent official claims of the complete eradication of illiteracy in Venezuela." In other words, we assert that the data rebuts the large program hypothesis, but cannot help us distinguish decisively between the small program and the no program hypothesis.

In this context, we believe that the relevant question to ask at this stage in the debate is whether RW's contribution helps us further distinguish between these hypotheses. We evaluate their points in this context.

Adequacy of the Households Survey data

RW argue that "a household survey with just one question about whether a family member can read cannot be expected to capture most of the effects of a large-scale literacy program such as *Misión Robinson.*" They support this assertion on the identification of several potential biases in the households survey, such as overestimation of reading skills, change in perceptions of adequate literacy standards, or different death rates by age groups over time.

These biases obviously exist, and we recognize them explicitly in OR. Several of our results are intended to deal with these problems. For example, our use of a restricted sample in which we exclude self-reported literacy answers in our Table 3 is intended to deal with self-serving biases in literacy reports. Likewise, our use of cohort analysis is intended to control for the effect on literacy of changes in death rates among cohorts. Yet it is true that these solutions are imperfect and test the limits of what can be done with the survey, thus reducing the possibility of obtaining a precise estimate of the program's effect.

However, none of RW's objections provide a strong rationale for explaining why the data would fail to pick up the effect of a massive literacy program. The reason is simple. Imagine that the Venezuelan government did indeed give reading and writing classes to 1.5 million Venezuelans. Surely, if these people already felt that they could read and write before the program, then they would have answered "yes" to the literacy question both before and after the end of the program, as suggested by RW. But the fact is that there were 1.1 million people who were claiming **not** to know how to read and write in early 2003, according to the survey. So one would have to believe that either (i) the program did not reach these people, or (ii) these people still claimed not to know how to read and write even after finishing the program.

The first possibility (a massive program not reaching a substantial fraction of those who claimed to be illiterate) is hard to believe and in itself would be an indictment of the program's effectiveness. We would have to believe that the Venezuelan government devoted the massive resources necessary to put upwards of a million persons in classrooms and yet ended up putting *exactly* the wrong million in the program - those who already felt that they knew how to read and write – while systematically excluding those who claimed to need the program the most. In order to get an idea of the dimension of the necessary assumptions, imagine that the program was composed in 90% of people who claimed to know how to read and write, with only the other 10% coming from the group that believed to be illiterate. Even in that case, taking the government's claims at face value, 150 thousand people who claimed to be illiterate before the start of the program would have enrolled in it. But this number is more than three times as high as the largest point estimates found in our study (and, as we will show, in RW's analysis).

The second possibility is almost as farfetched. It would imply that upwards of a million persons who claimed not to know how to read and write were enrolled in a seven-week program, showed their skills by composing a letter to President Chávez at the end of the program, received a certificate that indicated that they had passed the *Robinson* program, yet would still assert that they did not know how to read and write when asked by an interviewer. It would appear that in order for this to be true the program participants would have to be extremely skeptical that anything that they did while in the program even remotely resembled a literacy course. Again, if this were true, it would in itself constitute a striking demonstration of the program's failure.

Therefore, the biases pointed to by RW do not help rescue the large program hypothesis. They do give reasons why one may be inclined to believe in the small program hypothesis, but are inconsistent with the idea that the government put 1.5 million people into the seven-week literacy program.

Sensitivity of results to trend specification

Rosnick and Weisbrot claim that our results are sensitive to the choice of a cubic trend. They write that "for the most basic regressions in the paper, the results presented are very far from robust, and appear to be simply an artifact. Indeed, the cubic trend is clearly an outlier among a wide range of polynomials."

In OR, we present a vast number of specifications in order to evaluate the literacy program. These include (i) contemporaneous effect, (ii) lagged effect, (iii) cumulative effect (iv) break in trend (v) analysis by age subgroups (vi) analysis by national age cohorts (vii) state panel regressions (viii) cohort-state panel regressions, and (ix) use of information on trainers per capita.

Surprisingly, RW claim that they are conducting "a far more expansive search of models", but concentrate on analyzing different polynomial trends in the lagged specification (where the Robinson effect is measured with a lag of one semester). This is particularly surprising after they themselves have argued *against* this specification when they assert that "it is not obvious why *Robinson* would have a constant effect on literacy." Given this criticism, one would have expected them to focus in the cumulative specification (which indeed we adopt for our latter state-level regressions) rather than on a regression which they themselves claim is misspecified.

RW present a battery of pruned and partially pruned regressions of which the overwhelming majority display a positive, significant coefficient. They take the unusual approach of pruning strongly insignificant coefficients in order to conduct a robustness test. While pruning may be adequate to conduct searches for a correct specification, this is not what robustness analysis intends to do. Robustness analysis is meant to evaluate the sensitivity of a result to different regression specifications. While pruning recommends dropping a coefficient that is insiginificant because of collinearity, this is not typically done in robustness analysis, as collinearity is indicative of the incapacity of the data to separately identify the effect of the variable in question and is thus indicative of a coefficient's fragility. For this reason, modern approaches to robustness take a Bayesian approach (Hoeting et al., 1999, Sala-i-Martin et al., 2004) in evaluating the sensitivity of a result to all reasonable right-hand side specifications.

We show the effects of carrying out this analysis by evaluating the effect of the Robinson variable in different specifications with increasingly complex polynomial terms (up to a tenth-degree polynomial, as in RW) in Table R-1. We confirm the main thrust of RW's results: in the lagged specification, the overwhelming majority of coefficient estimates are positive, with many of them significantly so. For example, when we use the logit transform as the dependent variable, nine of ten coefficients are positive, with 6 of them significantly so.

Is this evidence that "the cubic trend is clearly an outlier among a wide range of polynomials?" This would only be true if we can show that the assertion applies to the broader set of specifications presented in our paper. In Table R-1, we also show the results of doing the same analysis for the contemporaneous, cumulative, and break in trend specification. In the logit transform specification, only seven of these additional thirty specifications are significantly positive (while three are significantly negative). Indeed, in the break in trend specification, there are as many positive as negative point estimates.

In the eighty regressions presented in Table R-1, 59 (73.8%) display positive point estimates, while 21 (26.2%) display negative point estimates. Less than one third are positive and significant, while about one-ninth are negative and significant. Most importantly, in 54 of the 80 regressions (67.5%) it is not significantly positive. The results shown in our paper with the cubic trend are thus representative of the results using different polynomial specifications, rather than the outlier that RW claim it to be.

Table R-1:Results of Sensitivity Analysis

	Positive	Significant (5%)	Negative	Significant (5%)
Dependent Variable: Literacy Rate				
Lagged	8	5	2	1
Contemporaneous	8	1	2	1
Cumulative	7	4	3	1
Beak in Trend	5	3	5	2
Dependent Variable: Logit Transform				
Lagged	9	6	1	1
Contemporaneous	9	0	1	1
Cumulative	8	3	2	1
Beak in Trend	5	4	5	1
Percent of specifications	73.8%	32.5%	26.3%	11.3%

<u>Statistical vs. Economic Significance</u>

Let us accept for the sake of argument that the lagged specification that RW have concentrated on is indeed the most appropriate one. The results shown by RW (and confirmed by us) indicate that in the case of the lagged effect the preponderance of polynomial specifications indicate positive, significant effects of *Robinson*. What implication does this have for evaluating the three competing hypotheses presented above? RW do not make an attempt to answer this question, as they center exclusively upon the statistical significance of the results, rather than their economic significance.

In Table R-2 we present the average program effects from RW's preferred lagged specification, using all possible polynomial specifications. As the table shows, the average effect in these regressions is an increase in 0.20% in the literacy rate in the specification where the untransformed rate is the dependent variable, and 0.25% in the one where the logit transform is the dependent variable. In numbers, this would imply between 34 and 42 thousand persons who benefited from the program. In other words, RW's preferred specification strongly supports the small program hypothesis and is inconsistent with the large program hypothesis.

Table R-2: Estimated Program Effect from Lagged Coefficient Specification

Order of Polynomial	Literacy Rate	Logit Transform
1	-2.13%	-0.53%
2	1.00%	0.37%
3	-0.03%	0.25%
4	0.96%	0.62%
5	0.61%	0.45%
6	0.33%	0.28%
7	0.32%	0.28%
8	0.46%	0.32%
9	0.28%	0.25%
10	0.17%	0.17%
Average Effect	0.20%	0.25%
Adult Population (1-03)	17098862	17098862
Effect of Program (in persons, using average)	33685	41968

Conclusions

Rosnick and Weisbrot pose a set of criticisms of the use of the Venezuelan households survey to evaluate the *Robinson* literacy program, and argue that the cubic polynomial specification that we chose in Ortega and Rodríguez (2008) is a clear outlier. However, we have shown that once the specification search is augmented to include alternative specifications other than the lagged effect, more than 2/3 of the coefficient estimates are either statistically insignificant or have the wrong sign. Thus the cubic trend, rather than being an outlier, is representative of the patterns found in general among different trend specifications.

More importantly, none of the reasons presented by Rosnick and Weisbrot constitute a convincing argument that a large literacy program involving upwards of one million persons existed in Venezuela. Such a program would have had to have a significant effect on the population that considered itself illiterate before the start of the program, and such an effect is not visible in the data.

Rather, Rosnick and Weisbrot's preferred specifications are consistent with the existence of a small to moderate sized program that affected between thirty and fifty thousand persons. The existence of such a program is also within the confidence intervals of the overwhelming majority of point estimates that we reported in Ortega and Rodríguez (2008). Rosnick and Weisbrot's analysis is thus a valuable contribution in rebutting the hypothesis defended by a number of authors (e.g., Weisbrot, 2008) of the existence of a massive literacy program involving upwards of one million persons in Venezuela.

References

Hoeting, Jennifer A.; Madigan, David; Raftery, Adrian E. and Volinsky, Chris T. “Bayesian Model Averaging: A Tutorial.” *Statistical Science*, November 1999, *14*(4), pp. 382–417.

Ortega, Daniel, and Francisco Rodríguez (2008) “Freed from Illiteracy? A Closer Look at Venezuela’s *Robinson* Campaign,”, *Economic Development and Cultural Change*, October 2008 (forthcoming).

Rosnick, David and Mark Weisbrot (2008) “‘Illiteracy’ Revisited: What Ortega and Rodríguez Read in the Household Survey” reproduced: CEPR.

Sala-i-Martin, Xavier, Gernot Doppelhoffer and Ronald Miller (2004) “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach” *American Economic Review* September, 812-835.

UNESCO. 2006. *Literacy for Life: Education for All Global Monitoring Report 2006.* Paris: United Nations Educational, Scientific and Cultural Organisation.

Weisbrot, Mark (2008) “Neoliberalism has failed,” *Americas Quarterly*, Fall 2007, pp. 20-22.