

Jorge M. Streb

Comments on Verónica Cabezas, José I. Cuesta and Francisco A. Gallego, “Effects of short-term tutoring on cognitive and non-cognitive skills: Evidence from a randomized evaluation in Chile”

Fifth Annual Meeting of the Impact Evaluation Network. LACEA

This interesting paper on an important issue addresses three questions:

- (i) Impact of late interventions
- (ii) Link between cognitive and non-cognitive skills
- (iii) Cost effectiveness.

I was especially interested in seeing the role of tutors that helped fourth-grade students read, since I also have a ten-year old daughter (my daughter usually does not want my help and prefers to read on her own, perhaps because I am a bad storyteller).

My comments will start with cost-effectiveness and work back from there.

1. The cost-effectiveness of the program should be assessed using the total impact of the program on reading comprehension, instead of picking out the most successful subsamples. This way a more meaningful comparison can be done with the other programs in Table 13.
2. The paper looks at the effects by pairs: high or low IVE, high or low SIMCE scores. Why not look at both combined? The case of high IVE and low SIMCE scores could be singled out, since these are the instances when the effects seem to be more pronounced. Besides, these were the kind of factors that led to single out these schools in the first place. I see the other distinctions (public or private subsidized schools, metropolitan region or region VIII) as secondary compared to these. If the program has a most significant impact for high IVE and low SIMCE scores, this may be an indication of non-linear effects.
3. Related to point 2, since all students are tested or treated, there is additional information on which pupils initially had low scores (lowest decile, quartile, or below median). Perhaps this can allow get some more information on how the program impacted on the most disadvantaged students, insofar as low grades are related to adverse socioeconomic factors. When the program works properly, does it help most the most disadvantaged students, or some intermediate group?
4. There is a lot of emphasis that the treatment group and the control group have similar characteristics for the whole sample. The paper should specify if the same holds for the subgroups on which the paper focuses.
5. Instead of describing in so much detail the econometric results found in Tables 6 and Tables 8 through 11, the paper could be much more concise, simply emphasizing any

significant econometric result, and perhaps dropping some of the subsamples that are not as important to interpret the results. The space saved on this could be devoted to explaining more of the substance of the experiment itself: how the “shared-reading instructional approach” works, the test itself, and the questionnaire on the Taste for Reading. The wording of the questionnaire can have a great effect on the responses, so one cannot be sure what the questionnaire is measuring until one sees the raw questions.

6. Since program implementation varied a lot between schools (Table 5), the authors carry out an instrumental variable estimation to see the effects of more sessions on reading comprehension. They could try a similar approach to see if higher tutor turnover had a deleterious effect on the program outcomes. That seems to be the reason why they interpret the results in the metropolitan region (Santiago) are worse than in region VIII.
7. The paper could mention that it controls for initial scores in the regressions.
8. The paper does not specify if the sessions were conducted during school hours or after school. Since, I am now told, they were conducted during school hours, this can help explain why some significantly negative results appear in relation to use of language (not to reading comprehension): the tutors could have made them lose time compared to their ordinary teachers.
9. The students may not have reacted positively to the material (for example, interest in reading and enjoyment of reading did not increase) because the material was not suited to their specific interests. No description is made of the choices the students had of the fictional and non-fictional material that was used. Perhaps a program with more options can work better. My daughter, for example, does not like many of the books I give her, and does not read them. But she started reading “Diary of a wimpy kid” after her 10-year old cousin talked to her about it, and then she read another one, and she now wants to go on reading the whole collection. Hence, it would be interesting to know how the intervention was carried out: did they have choice in what to read? The same tutors can work very differently with fixed or flexible reading material.
10. Depending on the implementation of the material, as well as the choice of tutors, the tutoring program can mean a lot of different things. The students could be asked if they are interested in the material that is being offered through the program (or in what part of the material interested them most). The students could also be asked about how they reacted to the tutors, if they liked them, if they bonded or not, which might be related to the turnover variables emphasized by the authors. This could provide important feedback from the fourth-grade kids that might help to improve the program in the future.