Reviewer’s report

Title: Study protocol for the evaluation of a community based early childhood education and development program in Indonesia: A pragmatic cluster randomized controlled trial with supplementary matched control group.

Version: 1 Date: 9 June 2013

Reviewer: William Barnett

Reviewer’s report:

Major Compulsory Revisions

1. Beginning with the first full sentence on page 6 through the next two full sentences, the text is insufficiently clear for readers not already familiar with early childhood services in Indonesia to understand the context for the study. Early childhood services and programs can refer to a wide range of programs not all of which provide education. From this section, the reader cannot know what is meant by "prior to pre-school". Please inform readers when pre-school begins. It is stated that 85% of the poorest children are not participating in any kind of program. However, the reader does not know what is meant by "program." Does this include health programs and informal child care or does it refer only to formal preschool education and care programs? In addition, it is unclear how "poorest" is defined. Does this mean those in poverty, in extreme poverty, or near poverty?

2. Several statements made in the text on ages 16 and 17 regarding sample retention and replacement of children from baseline with new children at midline follow-up do not appear to be consistent. At one point it is stated that 101+ children were not available for reinterview because they had moved. Then it is stated in the next paragraph that 101 children had moved and were reinterviewed. Later it is stated that the number of children who moved and were not reinterviewed is unknown. Perhaps this is what the "+" refers to, but it is unclear. Please clarify this matter. Moreover, it is essential that the study clearly reports the number of children who were replaced or that this figure is unknown and the follow-up sample sizes in each cluster if these are not identical to the baseline. Figure 1 does not currently help in this regard as I cannot reconcile the numbers in Figure 1 with the explanation in the text. Although the number of children who died is quite small, it would seem to me that they are properly included in an analysis as participants for whom there are no benefits at follow-up not as participants with unknown status at follow-up.

3. Statements about the length of delay in testing also appear to be inconsistent. Page 25 states that baseline data collection took place on average 6 months after implementation and midline 7 months after Batch 3 implementation. Page 26 states that effects on baseline child development will be ignored because 3 months of implementation is expected to have little impact. Page 30 states that actual difference in follow-up between baseline and midline was 14 months (I
assume this means assessment rather than treatment) which would be consistent with a 3 month initial delay and a 7 month midline delay. The formula for estimating treatment impact implies that delays are equal at baseline and midline. I do not know why the authors would assume that there is no impact of the delay as this is measurable. Comparison of Batch 1 and Batch 3 baseline should provide a suitable measure of the problem caused by the delay. Admittedly this may not be tested with as much power as desired, but it is better than simply assuming there is no difference.

4. The paper should discuss in the methods sections protections from false positives relating to multiple comparisons. Nowhere does the paper report the number of measures and what will be done to account the number of estimates of treatment effects conducted while taking into account interdependence of the outcome measures.

5. The paper should discuss potential heterogeneity of treatment effects and what will be done to investigate these. Based on past research, systematic differences should be expected based on what is actually implemented as a result of the program, which may differ greatly by type as well as quantity from one location to another, and by child and family characteristics. Children who are seriously malnourished at baseline and whose parents have the lowest levels of education might be predicted to benefit more (though the research is not uniform with respect to this issue, there are enough that point in these directions to make this of particular interest).

Minor Essential Revisions

1. In discussing the analyses to assess validity of the measures the paper omits looking at associations with age, this seems likely to be an unintentional oversight that should be corrected.

Discretionary Revisions

1. Given the nature of the program, the study analysts should consider developing measures of dosage. This will be complicated and could range from a simple measure of amount of funding actually spent on service to hours of direct services to multiple indicators of the amount of each of a several different services provided.

2. The discussion of the study recognizes that the sample cannot be assumed to be representative because of the way in which it was obtained. However, it would be useful to know how similar or dissimilar it is to Indonesia as a whole with respect to its measured characteristics and what part of the country it might represent. If the sample is sufficiently diverse, it may be possible to simulate results for a more representative sample from the study data. Two key aspects of representativeness might be discussed: local capacity to effectively implement the program; and, the demographics of the children and families served.

**Level of interest:** An article of outstanding merit and interest in its field
Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.