

Appendix A: The Ages and Stages Questionnaire (ASQ)

The Ages and Stages Questionnaire (third version, ASQ henceforth; Squires et al. 2009) was administered to measure child development in the children of the sample. The ASQ consists of age-specific questionnaires, a total of 10 for children between 6 and 24 months of age. While the ASQ was originally designed as a questionnaire to be completed by the mother, for this study it was administered by an interviewer to the child in the presence of his or her mother or main caregiver in the home.

The Spanish version of the ASQ was modified by local psychologists to adjust it to the Peruvian context. Moreover, in order to establish homogeneous administration protocols across the sample, the team of psychologists defined specific items that interviewers were required to administer directly to the child, while others could be collected by maternal report. As a result of this adjustment, across all age-specific questionnaires, the majority of items of the fine motor and problem-solving scales were administered directly to children, while the majority of items in the communication scale were collected by maternal report. More precisely, across all age-specific questionnaires, on average, only 7.5 percent of items in the fine motor scale and 10.6 percent of items in the problem-solving scale, but 60.1 percent in the communication scale, had to be collected primarily by maternal report. Careful piloting of the instruments ensured that the recommended administration protocol could be followed during data collection.

Each ASQ age-specific questionnaire includes six items per developmental area. Every item receives a score of 10 if the child completes it, 5 if she does it sometimes, and 0 if she does not do it, for a maximum of 60 points per scale. To reduce the number of children who reach the test ceiling, children who correctly completed all 6 items in a scale were also administered the first 3 items of the questionnaire corresponding to the next age group (as in Rubio-Codina et al. 2016). This adjustment allowed children who otherwise would have reached the test ceiling (10-11 percent of the sample, depending on the scale) to be tested on three additional, more difficult items. As a result of this adjustment, ASQ raw scores are over a total of 90 points.

The psychologists who adjusted the language and administration of ASQ also carried out its test-retest in a sample of 26 children, with an average difference of two weeks between the first and second administration. Considering 9 items per developmental area, the intra-child correlation (ICC) obtained was 0.80-0.85.

The ASQ was administered by a team of 23 interviewers, selected out of 36 who participated in the training. Interviewers had post-secondary education, but no prior experience administering the ASQ. They had all worked before as survey enumerators. Interviewer training was conducted by one of the psychologists who adapted the ASQ. It included 5 days of classroom training and a minimum of 10-15 practice administrations, some of them supervised by the trainer. All ASQ administrations were conducted in Spanish.

Table A1 shows the internal consistency of the instrument. The Cronbach's alphas were calculated separately for each questionnaire. The values reported refer to the average of the 10 alphas (i.e., one per questionnaire) for each developmental area. Cronbach's α for the ASQ were within a reasonable range for the three areas: $\alpha=0.644$ for communication, $\alpha=0.567$ for fine motor, $\alpha=0.592$ for problem solving, and good for the total (i.e., the items of all three areas together) with $\alpha=0.751$. The correlation coefficients are all high.

Table A1. Internal consistency of the ASQ-3

| | Correlation | Alpha |
|-----------------|--------------------|--------------|
| Communication | 0.703*** | 0.644 |
| Fine Motor | 0.763*** | 0.567 |
| Problem Solving | 0.753*** | 0.592 |
| Full Scale | | 0.751 |

Note: Cronbach's alphas and Pearson correlation coefficients between each developmental area and the total score. *** indicates significance at 1 percent.

Appendix B: The Toddler Classroom Assessment Scoring System (CLASS)

Process quality in the classroom was measured using the Toddler Classroom Assessment Scoring System CLASS-Toddler (CLASS hereafter; La Paro et al. 2012) coded from four 20-minute videos recorded over the course of a normal day at the center. The CLASS is an observation instrument for use with children 15-36 months of age. It assesses the quality of teacher-child interactions, in two domains: emotional and behavioral support, and engaged support for learning.

Each video segment was coded twice, by two different coders assigned at random. A team of six CLASS-certified observers did the coding. Coders carried out daily group exercises supervised by a CLASS-certified trainer to stay reliable over time. When scores differed by more than 2 points (1 point in scales of less variability), a third coding was carried out, and the score for the segment corresponds to the average of the two with the smallest discrepancy. Classroom scores are averaged for the two coders and four segments, with each domain receiving the same weight.

Table B1 presents the internal consistency of the instrument using two types of indicators: the correlation between each subscale and the overall score of the instrument, and Cronbach's alphas. Overall, the instrument shows excellent consistency with both indicators.^{1,2} The correlations are all in the very high range (i.e., above 0.80), with two exceptions: Regard for Child's Perspectives in the high range (ICC=0.73), and Negative Climate in the moderate range (ICC=0.53). The lower correlation for Negative Climate as compared to other subscales could be due to its low variability in the sample. Cronbach's α were good: $\alpha=0.87$ for Emotional and Behavioral Support, $\alpha=0.85$ for Engaged Support for Learning, and $\alpha=0.91$ for the full scale.

Table B1. Internal Consistency of CLASS

| | Correlation | Alpha |
|---|-----------------|--------------|
| Emotional and Behavioral Support | 0.917*** | 0.870 |
| 1. Positive Climate | 0.916*** | |
| 2. Negative Climate | 0.530*** | |
| 3. Teacher Sensitivity | 0.931*** | |
| 4. Regard for Child Perspectives | 0.734*** | |
| 5. Behavior Guidance | 0.846*** | |
| Engaged Support for Learning | 0.914*** | 0.853 |
| 6. Facilitation of Learning and Development | 0.881*** | |
| 7. Quality of Feedback | 0.852*** | |
| 8. Language Modeling | 0.897*** | |
| Full Scale | | 0.906 |

Note: Cronbach's alphas for CLASS dimensions and domains and Pearson correlation coefficients between CLASS dimensions and domains with the total score. All correlation coefficients are significant at 1 percent.

¹ Generally, the literature considers a measure of internal consistency to be reasonable when it falls within the range of 0.60 to 0.70 for both indicators.

² Correlations in the range of 0.80 to 1 are considered very high, high in the range of 0.60 to 0.80, moderate in the range of 0.40 to 0.60, low in the range of 0.20 to 0.40, and very low when less than 0.20.