

**Abstract Title Page**

*Not included in page count.*

**Title:** Measuring critical education processes and outcomes: Illustration from a cluster randomized trial in the Democratic Republic of the Congo

**Authors and Affiliations:**

Peter F. Halpin, Ph.D.  
Department of Humanities and Social Sciences  
Steinhardt School of Culture, Education and Human Development  
New York University  
Kimball Hall, Room 316E  
246 Greene Street  
New York, NY, 10003  
Tel: (212) 998-5197  
Email: peter.halpin@nyu.edu

Catalina Torrente, Ph.D.  
Yale Center for Emotional Intelligence  
Yale University  
340 Edwards Street  
New Haven, CT, 06520  
Tel: (614) 313-6262  
Email: catalina.torrente@yale.edu

## **Abstract Body**

*Limit 4 pages single-spaced.*

### **Background / Context:**

*Description of prior research and its intellectual context.*

Using reliable and valid measures of students' outcomes which are sensitive to change is critical for obtaining interpretable and therefore useful results from evaluations of school-based interventions. While measurement development for use in experimental evaluations receives a great deal of attention in the U.S., it lags behind in low-income countries.

The Early Grade Math Assessment (EGMA, RTI, 2009) and Early Grade Reading Assessment (EGRA, RTI, 2009b) were developed by RTI international to provide snapshots of students' emerging math and reading abilities in low-income countries. The assessments use a series of discrete sub-tests (e.g., number identification, quantity discrimination, phonemic awareness, familiar/unfamiliar word decoding) and are increasingly employed by governments and Non-Governmental Organizations (NGOs) to diagnose the status of children's abilities and to fine tune interventions to meet the academic needs of children. As the popularity of evidence-based practices grows, EGMA and EGRA are also being used in experimental evaluations of school-based interventions. However, the use of multiple sub-test scores may inflate the risk of capitalizing on chance and undermine power to detect intervention impacts.

In addition to measuring children's academic abilities, research points to children's learning environments and socio-emotional wellbeing as important predictors of learning and other important life outcomes (Durlak et al., 2011). Sensible instruments to measure children's wellbeing and the quality of learning environments in low-income and conflict-affected countries are scarce. Developing and testing measures that can capture these characteristics of children and their learning environments in reliable, face- and culturally-valid ways is critical to improving our strategies for the promotion of positive outcomes for all children.

### **Purpose / Objective / Research Question / Focus of Study:**

*Description of the focus of the research.*

The goals of the study are to describe and discuss our conceptual and analytical approaches to developing valid and internally consistent measures of social-emotional processes and academic outcomes for use in a cluster randomized trial of Opportunities for Equitable Access to Quality Basic Education (OPEQ) in the Democratic Republic of the Congo (DRC). First, we will present our approach to developing summary scores of children's early math and reading abilities. Second, we will present psychometric analyses used to refine, reduce, and examine the underlying structure of scales intended to measure children's perceptions of school quality and socio-emotional wellbeing.

### **Setting:**

*Description of the research location.*

Data for this study come from the first and second waves of a cluster randomized trial of OPEQ, an intervention developed by the International Rescue Committee which embeds socio-emotional learning principles and practices into high-quality math and reading lessons. OPEQ is being implemented in three eastern provinces of the DRC, a country where extreme poverty, chronic

divestment in education, and political unrest place children at high risk of academic underachievement and school dropout.

### **Population / Participants / Subjects:**

*Description of the participants in the study: who, how many, key features, or characteristics.*

Eighty-four schools from Katanga, one of the largest provinces in the DRC, participated in data collection for the current study. At baseline, schools had student populations ranging from 82 to 1,130 students. At each school, approximately eighty one students in grades 2–4 were randomly selected to participate in the evaluation. On average, children were 10.4 years-old (SD: 2.0); the vast majority (70%) reported primarily speaking Swahili –although French is the main language of instruction starting on 3<sup>rd</sup> grade—; 14% reported going to bed hungry “often” or “sometimes” in the month before the survey; and about 6% reported being “often” or always” absent from, or arriving late, to school.

### **Intervention / Program / Practice:**

*Description of the intervention, program, or practice, including details of administration and duration.*

OPEQ aims to enhance teachers’ motivation, the quality of school settings and teaching practices, and children’s academic achievement and socio-emotional wellbeing. The intervention has two primary and interrelated components. First, an innovative curriculum which integrates high quality reading and math lessons with IRC’s Healing Classrooms, a protocol of techniques to create safe and inclusive learning environments for all learners, is built into a teacher training package. Second, a school-based collaborative professional development system of continuous in-service teacher training and coaching is implemented. The structure is based on an historical practice of the DRC’s educational system: the Forums of Pedagogical Exchange (FPE). FPE’s consist of teacher-learning circles that are designed to meet: weekly at grade level; monthly at school level; and quarterly at school cluster (2–6 schools) level. FPEs enable teachers to collaboratively explore their practices, brainstorm solutions to challenges and identify and celebrate successes. These services are delivered by Master Trainers (MT; one per cluster of 2 to 6 schools) composed of teachers, headmasters, pedagogical advisors, inspectors and key technical staff from the Ministry of Education.

### **Research Design:**

*Description of the research design.*

A total of 203 schools nested in 54 clusters (i.e., groups of 2 to 6 geographically proximal schools) and 6 educational subdivisions (i.e., Kongolo, Kambove, Mutshatsha, Lubumbashi, Kasenga, and Kalemie) were selected to participate in the OPEQ intervention. By means of public lotteries carried out in each subdivision, all clusters were randomly assigned to different treatment cohorts. The Pilot Cohort (n= 20 clusters) started receiving the intervention in 2011; Cohort 1 (n =17) started in 2012; and Cohort 2 (n=17) started in 2013. The analyses in the current study use data from years one and two.

### **Data Collection and Analysis:**

*Description of the methods for collecting and analyzing data.*

*Data collection.* Out of 203 schools targeted by OPEQ, eighty-four were randomly selected for data collection. One school per cluster was selected if the cluster had three schools or less, and two schools were selected if the cluster had more than three schools. Children in second to fourth grades were randomly selected to participate in the evaluation. Sampling of children was stratified within schools and grade levels to ensure equal representation of students in different grades. Assent and consent were requested from all children at the time of data collection and refusal to participate was

very rare. The Ministry of Education and field team widely advertised the evaluation in each school and community to ensure that parents were fully informed and had the opportunity to ask any questions, raise any concerns, and opt out.

At all points of data collection, students were randomized to complete different sets of assessments with the aim of reducing participant burden. At baseline (wave 1), children were administered a demographic survey and two out of four assessments: EGMA, EGRA, or two versions of the socio-emotional wellbeing survey (SEW). In wave 2, students were administered the demographic and SEW survey (a single version was used after baseline), and were randomized to the EGMA or the EGRA. Child surveys and assessments were conducted in French (the official language of instruction) by local staff trained in data collection procedures by the OPEQ team. Swahili (the most common local language) was used for the demographic and SEW surveys, and to provide instructions for the EGMA when children had difficulties understanding French.

*EGMA and EGRA Analyses.* At baseline, the analysis sample included 2,938 children with EGMA and 2,951 children with EGRA data. Both assessments consist of a number of subscales (e.g., number identification, phonemic awareness), and the subscale scores are obtained as the number correct responses to the items within each subscale. For EGMA, most subscales had mild to moderate censoring from below, with an average of about 16% of the scores in the lowest category. The censoring was far more severe for the EGRA subscales, with an average of about 57% of responses in lowest category, and with no consistent distribution for the non-zero responses to the different subscales. Because the two assessments yielded very different data, two different approaches were taken for creating summary scores. For the EGMA, the subscales were treated as censored normal and an exploratory factor analyses (EFA) was conducted. For EGRA, the sub-tests were dichotomized into zero / non-zero responses and Item Response Theory (IRT) analyses were conducted on the dichotomized sub-tests. EFA and IRT analyses were implemented in Mplus version 6.12 (Muthen & Muthen, 1998-2011). Models were compared on the basis of their goodness of fit (RMSEA, TLI, and CFI), and a number of summary scores were obtained for each test.

The exploratory models from wave 1 will be tested for measurement invariance in wave 2. Measurement invariance allows for testing of progressively stronger hypotheses about the consistency of a scoring model in two or more groups. In order for the same summary score to be applicable to the two groups, the model parameters must be consistent over groups. When this is the case, any observed differences between the groups can be attributed to differences in their ability, rather than differences in performance of the assessment. Establishing measurement invariance for the EGRA and EGMA scales is therefore crucial for using the summary scores as a metric of the effectiveness of the OPEQ intervention.

*Socio-emotional wellbeing Analysis.* The baseline sample for these analyses consisted of children in second to fourth grades who were administered the demographic survey and both of the SEL questionnaires. The analysis sample was randomly split in half to reduce the risk of capitalizing on chance. Exploratory and confirmatory factor analyses were performed to examine the underlying structure of the scales utilized. Mplus version 5.2 (Muthen & Muthen, 1998-2011) and a WLSMV estimator were employed to adjust for the clustering of children and schools within the subdivisions, and to model categorical variables (students answered all questions using a 4-point Likert scale).

Supplementary internal reliability analyses were conducted in SPSS version 20. Confirmatory factor analyses will be performed with the second wave of data to examine the stability of the results.

### **Findings / Results:**

*Description of the main findings with specific details.*

*EGMA and EGRA.* Preliminary results for EGMA indicate that a two-factor solution (RMSEA = .039, CFI = .99; TLI = .97) was a better fit than a one-factor solution (RMSEA = 0.097 CFI = .91; TLI = .84). A promax rotation showed that the two factors were quite highly correlated ( $r = .74$ ), with one factor corresponding to addition/subtraction and the second to geometry. Three approaches were taken for obtaining factor scores, with an attempt to discover if computationally inexpensive unit-weighting schemes were comparable to more complicated methods (see Grice, 2001, for review). We compared unit weighting based on factor structure, unit weighting based on Thurstones's regression method; and the modal posterior estimates provided by Mplus. The first two methods led to equivalent weighting schemes, and these were highly correlated with the scores provided by Mplus. Due to the computation simplicity of the unit weighted scores, they can be preferred in the present application.

IRT analyses for EGRA suggest a good fit to the data (RMSEA = .066, CFI = .99, TLI = .98), and factor loadings suggest that not all sub-tests provide equally useful information regarding children's literacy. Reading comprehension and Oral passage reading had the highest discriminating power, followed by Familiar word reading and Unfamiliar word decoding. The sub-test with the least discriminating power was Initial sound identification, which was also the most difficult sub-test. For the IRT analysis, simple scoring methods were not applicable so only the modal posterior estimates of Mplus were considered. We are currently working on wave 2 of the analysis.

*Socio-emotional wellbeing.* Wave 1 results indicate that four internally consistent and face-valid constructs were supported by the data: children's perceptions of supportive schools and teachers (17 items,  $\alpha = .83$ ); children's perceptions of schools and classrooms as predictable and cooperative settings (10 items,  $\alpha = .85$ ); children's self-reports of victimization (5 items,  $\alpha = .77$ ); and mental health (12 items,  $\alpha = .83$ ). Analyses with wave 2 are currently underway.

### **Conclusions:**

*Description of conclusions, recommendations, and limitations based on findings.*

The analyses in this paper illustrate approaches to measuring children's academic outcomes, perceptions of school quality, and socio-emotional wellbeing, as part of an impact evaluation in the Democratic Republic of the Congo. As such, they contribute to advances in devising more reliable and valid instruments to inform the development of increasingly effective educational interventions in diverse parts of the world. These findings provide a model for future use of evaluative educational instruments in low-income countries, and confirm the importance of critically evaluating such instruments for use in new populations. The other two papers in this symposium will use these measures to estimate the impact of the OPEQ pilot year of intervention on the math and reading measures and the perception of school and social-emotional measures described above.

## Appendices

*Not included in page count.*

### Appendix A. References

*References are to be in APA version 6 format.*

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D. and Schellinger, K. B. (2011).

The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.

Grice., J. W. (2001). Computing and Evaluating Factor Scores. *Psychological methods*, 6, 430-540.

Muthen, L. K., & Muthen, B. O. (1998–2007). Mplus user's guide (Fifth Edition). Los Angeles, CA: Muthen & Muthen.

RTI International. (2009). Early grade mathematics assessment (EGMA): A conceptual framework based on mathematics skills development in children. Retrieved from website: [https://www.eddataglobal.org/documents/index.cfm/EGMA\\_ConceptualFramework\\_23Dec09\\_Final.pdf](https://www.eddataglobal.org/documents/index.cfm/EGMA_ConceptualFramework_23Dec09_Final.pdf)

RTI International. (2009b). Early grade reading assessment toolkit. Prepared for the World Bank. Research Triangle Park: RTI.