

The Psychometric Quality of the Preschool Child Observation Record: Does it Pass the Test for Use in Child Care Programs?

Katherine Barghaus (Scholar) and John Fantuzzo (Mentor)

University of Pennsylvania

Child Care Research Scholars (Grant# 90YE0138), 2012 – 2013

Project Description

The purpose of this study is to investigate the psychometric quality of the second edition of the Preschool Child Observation Record (COR-2; HighScope, 2003). The COR-2 is a teacher observational rating scale of early learning and development that consists of 32 items with five skill points each. The COR-2 is the second most widely used measure in the largest federally funded early childhood program for children from low-income households—Head Start. Despite its widespread use, there are no published studies on the psychometric quality of the COR-2. To bridge this gap, this research aims to provide comprehensive information on the validity and reliability of the COR-2. Such information can be used not only to evaluate the COR-2, but also to further its development. The validation approach created for this study is also advanced as a general model for evaluating and developing early childhood assessments.

Research Questions

To investigate the psychometric quality of the COR-2, this study evaluated it against components of quality assessment outlined in the *Standards for Educational and Psychological Testing* (the *Standards*; AERA et al., 1999). These include validity evidence based on an assessment's content, response process, internal structure, and relationships to other variables, as well as evidence of reliability and unbiasedness. Guided by the *Standards*, the following research questions were investigated:

- 1) *Content Validity*: To what extent is there validity evidence based on the content of the COR-2?
- 2) *Response Process Validity*: To what extent is there validity evidence based on the response process of the COR-2?
- 3) *Internal Structure and Reliability*: (a) Does the COR-2 capture six valid and reliable factors of development corresponding to the categories

listed in the testing manual (i.e., Initiative, Social Relations, Creative Representation, Movement and Music, Language and Literacy, and Mathematics and Science)?

(b) If not, what is the valid and reliable internal structure of the COR-2?

(c) Do the five skill points for each item correspond to a valid development sequence?

- 4) *Relationship to Other Variables*: Is there validity evidence based on the relationships of the COR-2 to other validated assessments?
- 5) *Temporal Invariance*: Is the internal structure of the COR-2 invariant across time so that growth on each of the factors can be captured?
- 6) *Unbiasedness*: Is the internal structure of the COR-2 invariant across sex, race, and English language learning status?

Sample. Two sources of data were employed to investigate the research questions. To examine the content and response process validity (research questions 1 and 2), information was collected from the materials in the COR-2 assessment kit and printed materials on the psychometric quality of the COR-2 and of the first edition of the measure. Based on the *Standards* and research literature on sources of content and response process validity, an interview guide was developed to collect additional information from the developers of the COR-2.

Quantitative data to investigate the COR-2's internal structure, reliability, external validity, temporal invariance, and unbiasedness (research questions 2-6) came from all preschool children in Head Start in the 2006-2007 school year in a large, poor Mid-Atlantic city. The analysis sample consisted of 4,071 children with COR-2 data for the fall, winter, and spring. The COR-2 was required by the district at the time to monitor children's learning and development. Approximately half the children in the sample were

male, 5% were Caucasian, 70% were African American, 18% were Latino, 3% were Asian, 4% were other, and 5% were English language learners. The average age of the children was approximately 3.5 years old. The study teachers were required by the district to have a bachelor's degree and certification in early childhood education. The two-day COR-2 training recommended by HighScope was provided to a subset of teachers who then trained the other teachers at their school who had not attended the training. In addition, informal follow-up training was provided in subsequent professional development sessions.

Method. Drawing on the *Standards*, psychometric research, and research on assessment development, this study created and applied a comprehensive, systematic model to examine the psychometric quality of the COR-2. The model includes research-based indicators of psychometric quality and methods to assess the quality of observational rating scales like the COR-2. The information collected from printed materials and interviews on the COR-2 was reviewed for evidence of each of the research-based indicators of content and response process validity. The internal validity of the COR-2 with respect to its dimensionality and the item skill point functioning was examined using exploratory and confirmatory factor analysis and item response theory. If evidence was found for a valid and reliable internal structure, the factors were assessed further in terms of their relationship to other variables using correlational analyses and invariance across time and important groups using structural equation modeling.

Progress Update

Preliminary analyses have been performed to address the research questions. Validity evidence based on the content and response process points to potentially fruitful avenues to pursue in order to further develop the COR-2. The factor analytic and item response theory investigations suggest several significant issues with the internal structure of the COR-2 with respect to its dimensionality and item functioning. However, these analyses are still preliminary because the overarching validation model is currently being

vetted by experts in the various methodologies employed. For example, a plan has been made to further unpack the qualitative data with the help of expert. These consultations are being sought to receive peer review prior to submitting to academic journals and widely disseminating the findings.

Implications for Policy/Practice

This study employs multiple quantitative and qualitative methodologies to provide a holistic evaluation of the COR-2 to inform its further development. More broadly, this research has developed and is validating a model for providing high-quality evidence on early childhood assessments. In doing so, this work responds to recent federal mandates for scientific evidence on the quality of assessment. For example, the current Race to the Top Early Learning Challenge requires states to use assessments with evidence of their validity, reliability, and unbiasedness. Thus, this research will help to provide the type of information currently most pertinent to early childhood policy makers, administrators, and educators.

Implications for Research

Scientifically based measures are created through an iterative cycle of rigorous development and validation. To help guide this cycle, this research has developed and is validating a comprehensive, systematic model of evaluating the psychometric quality of assessments. The aim of the model is to provide an explicit, disciplined, and generalizable approach to assessment validation research. Therefore, this work will not only inform future research on the COR-2, it will also help to guide research on other early childhood assessments.

Contact

Katherine Barghaus
Penn Child Research Center
3701 Locust Walk, Room C-2
Philadelphia, PA 19104
Phone: (203) 313 – 0220
Email: barghaus@gse.upenn.edu
