Measurement Equivalence of Teacher Rating Scales of Social-Emotional Development with a Culturally and Linguistically Diverse Sample of Preschool Children

Aaron Douglas$^2$ and Sally Atkins-Burnett$^2$

BACKGROUND AND RESEARCH QUESTION

The cultural and linguistic backgrounds of the children served by Los Angeles Universal Preschool (LAUP) programs differ from the nationally representative sample used as the normative group for measures of social and emotional development.

Problems can arise for the validity of scores for these children because (1) they are from linguistic and cultural groups that are not adequately represented in the norming sample, and (2) the assessments may function differently for such children than for children from the reference group, after controlling for ability.

This poster will address the second of the two issues by posing the following question:

Is there measurement invariance on measures of social and emotional development and problem behaviors for children from different language groups, after controlling for ability?

METHODOLOGY

Structural equation modeling (SEM) was used to fit mixture latent trait models with two known latent classes of Spanish and English speaking children. This approach was used to investigate the presence of differential item functioning (DIF) on the following measures:

- Approaches to Learning from the Early Childhood Longitudinal Study-Kindergarten Cohort 1998-99 (ECLS-K) – 6 items
- Social Skills Rating System (SSRS; Gresham and Elliott, 1990) – Problem Behaviors – 10 items, Externalizing and Internalizing scales (The frequency scale was adapted to a 4-point scale to match the other two rating scales)
- Preschool and Kindergarten Behavior Scales, Second Edition (PKBS-2; Merrell, 2002) – three social scales, consisting of 34 items total
- PKBS-2 Social Skills
  - Standardized item alpha = .95 Social Cooperation, .93 Social Interaction, and .89 Social Independence.
  - One item per scale showed a difference of the factor loading between groups (Items 1, 2, and 20 for Social Independence, Cooperation, and Interaction, respectively).
  - One Social Interaction item (Item14) and one Social Independence item (Item 26) showed variant levels of difficulty – favoring children who spoke primarily or only English.

FINDINGS

No differences in the levels of difficulty for the categories of items by linguistic group.

One Social Interaction item (Item14) and one Social Independence item (Item 26) showed variant levels of difficulty – favoring children who spoke primarily or only English.

Strengths of relationships between dimensions differed between groups.

The relationship between the Externalizing and Internalizing dimensions was significantly stronger for the English speaking children compared to Spanish speaking children.

CATEGORIES OF ITEMS BY LINGUISTIC GROUP

- No differences in the levels of difficulty for the categories of items.
- Preschool and Kindergarten Behavior Scales, Second Edition (PKBS-2; Merrell, 2002) – three social scales, consisting of 34 items total

SAMPLE

- Fall 2007 UPCOS sample language groups 1, 5, 6, and 2, 3 (see poster on language routing for description) comprised of children who speak only or primarily English (N=856) and children who speak only or primarily Spanish (N= 633). Proficiency in English is being used as a proxy for acculturation to majority U.S. culture.
- Each teacher rated an average of 11.6 children.

METHOD

- The Huber-White option for variance estimation was used to correct for the clustering of children within classrooms when computing robust standard errors for parameter estimates of the models.
- A Bonferroni adjustment was implemented to ensure a family-wise level of significance at .05 for two-sample t-tests between estimates of model parameters, respectively, for compared groups.
- A Bonferroni adjustment was implemented to ensure a family-wise level of significance at .05 for two-sample t-tests between estimates of model parameters, respectively, for compared groups.

CONCLUSIONS

These analyses suggest that there is some evidence that these teacher ratings function differently in relation to the child’s cultural and linguistic background, after controlling for ability.

- Scores have different meaning for the two linguistic groups, and the reliability of the scores will differ partly due to DIF.
- The findings suggest that the measurement of social interaction is confounded by language. Both items showing differences in difficulty on the PKBS-2 are dependent upon language use (e.g., participating in discussions).


2Mathematica Policy Research, Inc.