CULTURAL EQUIVALENCE OF THE ESP: AN EXAMINATION OF DIFFERENTIAL ITEM FUNCTIONING WITH AMERICAN INDIAN CHILDREN

Hollie Hix-Small, [1,2] Jason W. Small, [1] Edward G. Feil, [1] and Paul Yovanoff [2]
[1]Oregon Research Institute, Eugene, OR; [2]University of Oregon, Eugene, OR

Oregon Research Institute

Introduction

American Indian communities often differ from the larger population with respect to social, spiritual, political and economic values and practices. Despite the interest within American Indian communities to improve their capacity to deliver early childhood education and screening services, the appropriateness of using screening instruments normed and validated on primarily Caucasian samples raises questions regarding the generalizability and interpretability of results. This study uses an Item Response Theory (IRT) differential item function analysis (DIF) to identify ESP item differences between an American Indian and Caucasian preschool sample.

Research Questions

1. Do ESP items function differently for Caucasian and American Indian preschoolers?

2. If DIF is present, does it represent nuisance or bias? In other words, do differentially functioning items represent the presence of an additional relevant dimension interpretable as item bias?

Why Culture?

▪Acceptability and non-acceptability of certain behaviors are based on cultural norms.

▪Parents and children are cognitively, linguistically, and emotionally connected to the language and culture of their home (NAEYC, 1996).

▪Parents interpret, respond to, and shape child behavior in accordance with culturally prescribed expectations and socialization goals.

▪Researchers have found that more Individualistic cultures value assertive and independent behaviors. In contrast, more collectivistic cultures value socially unobtrusive (e.g., restrained) and compliant behaviors that maintain social harmony.

Methods

Sample

▪110 Caucasian children, attending Head Start classrooms in the Pacific Northwest, ranging in age from 3-5 years (M = 4.39 years; SD = 0.60) and 39% of the participants were female.

▪69 American Indian children, attending Head Start and living on a reservation in the Western United States, ranging in age from 3-5 years (M = 4.21 years; SD = 0.62) and 30% were female.

Early Screening Project (Feil, Severson, & Walker, 1998)

▪Teacher-complete measure used to screen and identify children, aged 3-5, experiencing preschool adjustment problems.

▪8-item Social Interaction Scale measuring positive social behaviors.

▪16-item Critical Events Index assessing high intensity, low frequency behavior problems.

▪9-item Aggressive Behavior Scale measuring the frequency of aggressive behaviors.

▪8-item Adaptive & 9-item Maladaptive behavior indices assessing the student's teacher-related and peer-to-peer behavioral adjustment.

▪Normative sample (N = 2,853) was 69% Caucasian (as reported by their teachers), 16% Hispanic, 12% African American, & 3% American Indian or Asian.

Mean ESP scores: American Indian versus Caucasian

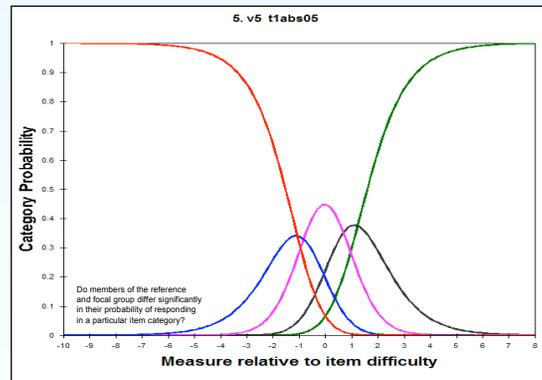| Scale	Total Sample M (SD)	American Indian M (SD)	Caucasian M (SD)
Adaptive Behavior (ABI)	26.94 (6.82)	26.49 (6.56)	27.22 (7.00)
Aggressive Behavior (ABS)	16.43 (7.66)	15.35 (7.03)	17.11 (7.99)
Critical Events (CEI)	1.08 (1.47)	0.55 (0.96)	1.42 (1.63)
Maladaptive Behavior (MBI)	22.23 (8.37)	20.71 (7.57)	23.19 (8.76)
Social Interaction (SIS)	33.98 (12.22)	31.77 (11.78)	35.38 (12.35)

IRT and DIF

▪Steps for detecting Differential Item Functioning using IRT or Rasch methodology:

▪Estimate item parameters for reference and focal groups.
▪Link separately estimated item parameters on the same scale (Vale, 1986).
▪If DIF is not present, parameters do not differ significantly.
▪If DIF is present, individuals from the examined groups *who are equal in level on the latent trait* do not have the same probability of endorsing an item.

DIF and the ESP

▪Dichotomous scales tested using the One Parameter Logistic model (1PL).
▪Polytomous scales tested using the Partial Credit model (PCM; Masters, 1982).
▪Models tested using WINSTEPS (Linacre, 1991-2005).

Sample category response curve (CRC) under the PCM

5. v5 t1abs05

Do members of the reference and focal group differ significantly in their probability of responding in a particular item category?

Results

Item Fit & DIF by ESP scale

Scale	Separation	Item Reliability	Outfit MNSQ(ZSTD)	Number of Items	DIF N (%)
ABI	3.12	.91	1.00 (-.20)	8	1(13%)
ABS	5.93	.97	.95 (-.20)	9	1(11%)
CEI	2.05	.81	.97 (.10)	16	1(6%)
MBI	3.41	.92	.98 (-.30)	9	1(11%)
SIS	5.85	.97	.97 (-.10)	8	0(0%)
Total	--	--	--	50	4(8%)
*Total after adjustment**	--	--	--	50	1(2%)

*Significant after Bonferroni adjustment at p < .05

▪Separation measures the spread of items across the latent trait continuum.
▪Expected value of mean square (MNSQ) is 1.0.
▪Expected value of standardized fit (ZSTD) is 0.0.

ESP items exhibiting Differential Item Functioning

Items	Scale	t-value
Expresses anger appropriately (reacts to situations without becoming violent or destructive).	ABI	-2.38
Makes lewd or obscene gestures.	ABS	3.30*
Has severely restricted activity level.	CEI	2.01
Responds inappropriately when other children try to interact socially with her/him.	MBI	2.79

Note: The *t* is the DIF contrast divided by the joint S.E. of the two DIF measures. It is equivalent to the Mantel-Haenszel significance test (WINSTEPS Manual, 2005). Group: 1 = Caucasian, 2 = American Indian.

▪Results suggest that teachers reporting on American Indian preschoolers have a lower threshold for endorsing problem behavior and, in turn, a higher probability or likelihood of endorsing the four ESP items described above.

▪ESP scales exhibit minimal significant DIF.

Limitations

▪Small sample size limits model complexity.

Conclusions

▪When adapting an instrument for use in different languages or cultures, researchers should investigate potential construct, method, and item bias (Sireci, 2005). IRT or Rasch-based DIF analysis affords researchers a tool for evaluating measurement equivalence and testing, at an item level, for differential item functioning between cultural groups.

Implications and Recommendations

▪Misapplication of assessment instruments with culturally-diverse populations may invalidate findings and preclude comparability with normative samples.

▪Use caution when borrowing a test from one culture and adopting it in another (Merenda, 2005).

▪Analyze instruments for potential construct, method, and item bias.

▪Assessment and interpretation of cross-cultural differences should be extended beyond test adaptation to all parts of the assessment process (Hambleton, 2005).