2014-12-17

# An Empirical Examination of the DECA Behavioral Concerns Subscale: Establishing Measurement Invariance Between Language Forms

Veronica A. Fernandez
*University of Miami,* vfernandez17@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

UNIVERSITY OF MIAMI


AN EMPIRICAL EXAMINATION OF THE DECA BEHAVIORAL CONCERNS
SUBSCALE: ESTABLISHING MEASUREMENT INVARIANCE BETWEEN
LANGUAGE FORMS


By

Veronica A. Fernandez


A THESIS


Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science


Coral Gables, Florida

December 2014

UNIVERSITY OF MIAMI


A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science


AN EMPIRICAL EXAMINATION OF THE DECA BEHAVIORAL CONCERNS
SUBSCALE: ESTABLISHING MEASUREMENT INVARIANCE BETWEEN
LANGUAGE FORMS


Veronica A. Fernandez


Approved:


_____
Rebecca J. Bulotsky-Shearer, Ph.D.
Associate Professor of Psychology

_____
Daryl B. Greenfield, Ph.D.
Professor of Psychology


_____
Randall Penfield, Ph.D.
Professor of Educational
and Psychological Studies

_____
M. Brian Blake, Ph.D.
Dean of the Graduate School

FERNANDEZ, VERONICA A.                          (M.S., Psychology)

An Empirical Examination of the DECA                (December 2014)
Behavioral Concerns Subscale: Establishing
Measurement Invariance Between Language Forms

Abstract of a thesis at the University of Miami.

Thesis supervised by Associate Professor Rebecca J. Bulotsky-Shearer.
No. of pages in text. (36)

Appropriate assessment instruments are necessary to accurately identify

children's behavior problems and to develop interventions within the classroom. Due to

increases in the cultural and linguistic diversity among the early childhood professionals,

especially Spanish-speaking teachers, measures are needed in teachers' dominant

language. In response to such a need, the publishers of the Devereux Early Childhood

Assessment (DECA; LeBuffe & Naglieri, 1999), a widely-used measure of children's

classroom behavior, have developed a Spanish-language form of the measure (DECA-

Spanish; LeBuffe & Naglieri, 2001). In the present study, the measurement properties of

the Behavioral Concerns subscale of the DECA were examined to confirm the invariance

between the two language forms and to evaluate the appropriateness of its use within a

large sample of diverse, low-income preschool children. Using a multiple group

confirmatory factor analysis, the two language forms were tested for configural and

metric invariance. Due to inadequate model fit to the  published factor structure, a new

two-factor structure, that distinguished between items that refer to Externalizing versus

Internalizing Behavior Concerns, was derived by means of exploratory factor analysis.

Multiple group comparison models using two-factor structure resulted in adequate fit

providing evidence that with this structure the English- and Spanish-language forms of

the Behavioral Concerns subscale of the DECA were invariant. Due to the inability to

confirm the factor structure, results from the DECA, using the published factor structure

should be interpreted with caution, especially when used within a low-income diverse

sample.

**TABLE OF CONTENTS**

# LIST OF TABLES

# LIST OF FIGURES

Page

Figure

**Chapter 1: Introduction**

High quality preschool programs, such as Head Start, have the opportunity to identify behavior problems early and provide comprehensive intervention services. Logically, appropriate assessment instruments are necessary in order to identify behavior problems within the classroom (National Association for the Education of Young Children [NAEYC], 2003; United States Department of Health and Human Services [USDHHS], 2007). In addition, as the early childhood workforce becomes increasingly culturally and linguistically diverse (Family and Child Experiences Survey [FACES], 2006), screening tools that can identify classroom behavior problems are needed in teachers' dominant language. The publishers of the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999), a commonly used measure of preschool children's classroom behavior, have developed a Spanish-language form of the measure. However, to date, there has only been one preliminary study by the publisher that compared scores on the English- and Spanish-language forms, and no empirical study has examined the psychometric properties of the Spanish-language form. In the presented study, the Behavioral Concerns subscale was empirically evaluated and the invariance between the English- and Spanish-language forms of this subscale examined, in a sample of low-income minority children at increased risk for exhibiting behavior problems within the preschool classroom.

**Behavior Problems within the Preschool Classroom**

There are increased national concerns regarding the social emotional needs of children from low-income families served within early childhood educational programs (Copple & Bredekamp, 2009; NCES, 2004; USDHHS, 2007). Behavior problems are

quite prevalent in preschool-aged children, with epidemiological studies suggesting that about 20% children experience difficulties (Lavingne et al., 1996). Children living in poverty are at increased risk of exhibiting behavior problems within the classroom (Cooper et al., 2008; Sinclair, 1993). Recent studies suggest that in programs serving low-income children, such as Head Start, between 30 and 41% of children exhibit significant behavioral problems (Barbarin, 2007; Feil et al., 2005; Qi & Kaiser, 2003; Webster-Stratton & Hammond, 1998). Since research suggests the negative and often long-term consequences of preschool behavior problems on academic and social outcomes (e.g., Denham, 2006; Raver, 2002), there has been an increased focus on identifying early behavior problems. Early identification of behavior problems can inform intervention efforts within preschool programs to promote more adaptive behavior that can support learning within the classroom context (Campbell, 2002).

Understandably, it is important to identify at-risk children at a very early age and provide interventions that can mitigate the potential negative effects of poverty and minority status on children's formative early learning experiences. Enriching early childhood experiences, such as participation in high quality preschool programs like Head Start, have been found to improve the developmental outcomes of low-income children (e.g., Campbell, Pungello, Miller-Johnson, Burchinal & Ramey, 2001; Reynolds & Ou, 2003). However, in order for such programs to attain their goal of promoting school readiness (NEGP, 1997; USDHHS, 2006), valid, reliable and sensitive measures are necessary to identify children's social and emotional needs and to inform instructional interventions within the classroom (Copple & Bredekamp, 2009; USDHHS, 2007).

**Culturally and Linguistically Diverse Preschool Teachers**

In addition to the increasing diversity of children and families enrolled in early childhood programs, recent statistics document the increasing diversity of the early childhood professional workforce. Teacher demographic reports from national Head Start studies, such as FACES, suggest that teachers are diverse with respect to their cultural and linguistic backgrounds. According to the most recent FACES (2006) data, about 60% of the Head Start teachers nationally identify themselves as racial or ethnic minorities. In the local Miami-Dade County program, the current Program Information Report indicates that about 50% of teachers speak at least some Spanish. In order for programmatic assessment practices to inform appropriate and timely early identification and intervention efforts, it is necessary to provide teachers that are Spanish-language dominant with valid and reliable Spanish-translations of early childhood assessments. Ideally, these assessments would be linguistically comparable and psychometrically equivalent with the original English language form.

**The Devereux Early Childhood Assessment**

The Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999) is a nationally standardized, norm-referenced teacher and parent behavior rating scale of children's social emotional adjustment. It includes three empirically derived subscales that assess 'Protective Factors' (Initiative, Self-control, and Attachment) and one subscale that assesses 'Behavioral Concerns.' The DECA is used by teachers in numerous Head Start programs across the U.S. as a programmatic behavioral screener to

inform mental health referral and intervention practices. In response to the need for Spanish-language measures, the DECA was recently translated into Spanish (DECA-Spanish; LeBuffe & Naglieri, 2001).

**Psychometric Evidence for the English Language Form of the DECA**

There is some evidence to support the reliability and validity of the English-language form of the DECA. Several studies, including the one conducted by the publisher, provide evidence for the construct validity of the Protective Factors. LeBuffe and Nalgeri's (1999) initial exploratory factor analyses for the protective items identified the three published subscales: Initiative, Self-control, and Attachment. Recent studies have replicated the published factor structure of the Protective Factor, with the same three subscales in a predominantly white sample (Jaberg, Dixon & Weis, 2009) and in a sample of low income Head Start preschoolers (Lien, & Carlson, 2009).

**Limitations of the Current DECA Research**

The extant literature is limited because it only provides initial support for the construct validity for the scores generated by the Protective Factor and its subscales, and not for the Behavioral Concerns subscale. According to the published manual, the Behavioral Concerns subscale was not derived by means of factor analysis, but rather, the ten subscale items were selected based on "their psychometric properties and their representation of a wide range of challenging behaviors" from among the 77 problem behaviors in the initial pool of items (LeBuffe & Naglieri, 1999, p.12). To date, all factor analytic studies of the DECA have excluded the Behavioral Concerns subscale and no study has examined the reliability and validity of the Behavioral Concerns subscale for low-income preschool children. Additionally, research indicates that the Behavioral

Concerns subscale mean scores are higher in low-income diverse samples than in the standardization sample (Brinkman et al., 2007, Lien & Carlson, 2009) which was comprised predominantly of white (76.3%) and few African Americans or Hispanic children (18.8% and 10.7%, respectively) (LeBuffe & Naglieri, 1999). Previous research has revealed the difficulty of confirming the structure of published measures when the measure was not developed or standardized for use with low-income populations (Fantuzzo, Hightower, Grim, Montes, 2002; Fantuzzo, Manz, McDermott, 1998; Konold, Hamre & Pianta 2003; LeBoeuf, Fantuzzo, & Lopez, 2010). Therefore, further research is needed if the Behavioral Concerns subscale is to be used in early childhood programs serving more ethnically diverse populations of low-income children.

**Development of the Spanish Language Form of the DECA**

According to the publisher, the following steps were taken to develop and validate the Spanish DECA form. (DECA-Spanish; LeBuffe & Naglieri, 2001)  The DECA was translated into Spanish and back translated into English. Then a small sample (of 44 bilingual parents & 48 bilingual teachers) rated children using both the English- and Spanish-language forms.  Paired sample *t*-tests indicated that there was no significant difference between the mean ratings across the two forms. Therefore, the publishers concluded that DECA ratings were comparable across language forms and that it was acceptable to use the English norms for the Spanish language version (LeBuffe & Naglieri, 2001).

There are some limitations to this approach. Smith, McCarthy, & Anderson (2000) stress the importance of empirically evaluating the dimensionality of any adapted measure. They caution against treating measures as equivalent or applying the norms of

the original measure to the new form, without first examining the measure's underlying factor structure. Aside from the small study by the publisher (LeBuffe & Naglieri, 2001) there is currently no other empirical evaluation of the equivalence between the English- and Spanish-language forms of the DECA. Additionally, no study has examined the published factor structure of the Spanish-language form of the DECA. Only the factor structure of the English-language form of the DECA Protective Factors has been examined empirically.

**Best Practice Recommendations for Translated Measures**

The *Standards for Educational and Psychological Testing,* which serve as national guidelines for the evaluation of tests and testing practices, delineate the following standards for the development and use of culturally sensitive and translated measures (Standards; American Educational Research Association [AERA], American Psychological Association [APA] & the National Council on Measurement in Education [NCME], 1999):

> 9.7 When a test is translated from one language to another, the methods used in establishing the adequacy of the translation should be described and empirical and logical evidence should be provided for score reliability and the validity of the translated test…
>
> 9.9 When multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability.

These standards suggest that prior to a translated measure's use for screening or early identification purposes, rigorous analyses of the measure's underlying factor structure, reliability, and invariance with the original language form should be established. Ignoring

these important considerations during the development of a second language form of an assessment instrument could result in cross-cultural inconsistencies, which can negatively influence the validity of the translated measure (Pena, 2007; Smith, McCarthy, & Anderson, 2000). In order to comply with such national standards, additional psychometric evidence for the Spanish-language form of the DECA is needed.

While best practices in cross-cultural measurement development stress the importance of establishing the equivalence of measures prior to and during the translation of an instrument, such processes are complex and are typically not considered. Often, the psychometric properties of an original measure's factor structure are simply applied to the translated language form (Deyo, 1984). This is not appropriate, because even when a high level of reliability is attained for scales on both language forms, construct validity could be quite low (Deyo, 1984). In other words, even when a translated form's scales demonstrate adequate internal consistency, they may not replicate the underlying factor structure of the translated form or be directly comparable to the original language form. In order to determine whether the Spanish-language form of the DECA is reliable and valid for low-income populations, and comparable to the English-language form, careful examination of both the construct validity and measurement invariance of the Spanish-language form of the DECA is needed.

**Measurement invariance.** Once construct validity is established, in order to make meaningful comparison between the scores derived from two language forms of an instrument, measurement equivalence is necessary; therefore, it is critical that an analysis of invariance is performed once a measure has been translated (Bollen, 1989; Drasgow, 1984; Horn, 1991; Vandenberg & Self, 1993). Measurement invariance examines the

degree to which an instrument yields the same distribution of scores for individuals who have the same value of a latent construct across different conditions (Horn & McArdle, 1992; Jöreskog, 1971; Meredith, 1993). For the purpose of the present study, the different conditions are the English- and Spanish- language forms of the DECA.

Specific statistical invariance tests are performed in order to determine if latent construct scores have the same meaning under different conditions (Horn & McArdle, 1992; Jöreskog, 1971; Meade & Lautenschlager, 2004; Meredith, 1993; Vandenberg & Lance, 2000). There are two statistical analyses that are used to examine empirically the equivalence between two forms of a measure: an analysis of *configural invariance*, which examines the basic fit of the factor structure for the forms of a measure; and an analysis of *metric invariance*, which examines whether the latent construct explains the indicators in the same way across groups (Horn & McArdle, 1992).

The present study examined the Behavioral Concerns subscale of the DECA (LeBuffe & Naglieri, 1999) as the first step of a larger project that aims to evaluate the measurement properties of the overall measure and its effectiveness in accurately identifying low-income minority children with emotional and behavioral concerns. The current study addressed two major research questions:

(1) Is the published factor structure for the Behavioral Concerns subscale of the DECA appropriate for a sample of diverse, low-income preschool children?

(2) Is the factor structure of the Behavioral Concerns subscale invariant across English- and Spanish-language forms of the DECA?

Due to the lack of rigorous psychometric support of the Behavioral Concerns subscale and prior research highlighting the difficulty in replicating the published factor

structures of behavior rating scales in low-income diverse sample (e.g., Fantuzzo, Hightower, Grim, Montes, 2002), it was expected that the published factor structure would not be appropriate for this sample of Head Start preschoolers. Furthermore, it was expected that measurement invariance would not be established between the English- and Spanish-language forms of the Behavioral Concerns subscale of the DECA, since translated measures are very infrequently found to be equivalent across languages (Schmitt and Kuljanin, 2008), especially when the factor structure of the translated form is not independently established (Deyo, 1984).

**Chapter 2: Method**

**Procedure**

Approval was obtained from the University of Miami's Institutional Review Board (IRB) for this study as part of a larger University- Head Start collaborative research partnership project involving the integration of several large administrative databases. These were programmatically collected by the Head Start program to meet the federal Head Start Performance Standards reporting and assessment requirements (USDHHS, 2007): (1) a child and family information database which included demographic and placement information (date of birth, gender, ethnicity, special needs status, center name, classroom assignment etc.) and (2) scores from the DECA, the program-wide classroom behavior screener. As there was no unique identifier across any of the databases Microsoft Integrated Services was employed and a probabilistic linking program was used, which included child first name, last name, date of birth, gender, and race/ethnicity combinations. Once the databases were integrated, the data were de-identified to protect participant confidentiality.

**Participants**

The sample included all children enrolled in the Miami Dade County Community Action Agency Head Start program in 318 classrooms, across 78 centers, for which the DECA was completed by lead teachers in the fall of 2008 ($N = 6,089$). Approximately 53% of the sample was female and children's ages in the fall ranged from 33 to 59 months ($M = 47.9$, $SD = 6.9$). Children were predominantly African American and Hispanic (55.4% and 43.8%, respectively), with .8% identified as being of another

ethnicity (including White/Non-Hispanic, Asian, Other, or Native Islander). Children in this sample were from families that met the federal income requirement for enrollment in Head Start indicating a sample of low-income children.

Participants comprised two groups based on whether the teacher completed the DECA in English or Spanish. There were 5,197 DECA completed in English and 860 completed in Spanish (32 records did not have an indication of the language in which they were completed and were deleted from the file for data analyses).

**English-language DECA group.** Of the 5,197 children for whom teachers completed the English-language form of the DECA, approximately 52% were female and their ages in the fall ranged from 33 to 59 months ($M = 48.1$, $SD = 6.9$). Children were predominantly African American and Hispanic (60.9% and 38.3%, respectively), with 0.8% identified as being of another ethnicity (including White/Non-Hispanic, Asian, Other, Native Islander).

**Spanish-language DECA group.** Of the 860 children for whom teachers completed the Spanish-language form of the DECA, approximately 53% of the children were female and their ages in the fall ranged from 34 to 59 months ($M = 47.1$, $SD = 6.9$). Children were predominantly Hispanic and African American (77.8% and 21.4%, respectively), with 0.8% identified as being of another ethnicity (including White/Non-Hispanic, Asian, Other, Native Islander).

Miami-Dade Head Start program demographic records indicated that teachers in the overall program were predominately African American and White (49.7% and 35.8%, respectively), with 14.4% identified as being of another ethnicity (including Asian, Bi-racial and unspecified). Approximately half of the teachers identified themselves as

Hispanic. There were approximately 44% of teachers with a Bachelor's degree, 22% with an Associate's degree, and 14% with a Child Development Associate credential.

**Measures**

**Classroom behavioral concerns.** The Miami-Dade Head Start program uses the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999; DECA-Spanish; LeBuffe & Naglieri, 2001) to assess children's classroom behavior. The DECA is a standardized preschool measure that includes three empirically supported subscales that assess 'Protective factors' (Initiative, Self-control, and Attachment) and one subscale that assesses 'Behavioral Concerns.' Teachers are asked to use a five-point Likert scale (0 = never, 1 = rarely, 2 = occasionally, 3 = frequently, and 4 = very frequently) to rate how often each student displayed the specified behaviors over the previous four week period. The DECA provides standardized $T$ scores for each subscale derived from the standardization sample. LeBuffe and Naglieri (1999) report the internal consistency reliability for teachers to be .94 for the total Protective Factor and .80 for Behavioral Concerns subscale. Test-retest reliability coefficients for teachers are reported as .94 for the total Protective Factor and .68 for the Behavioral Concerns subscale. For the present study, the DECA was filled out by lead teachers in the fall of the Head Start school year in their language of preference (English: $n$= 5,197 or Spanish: $n$= 860).

**Data Analytic Plan**

Structural equation modeling (SEM) was employed in M*plus* version 6.10 (Muthén & Muthén, 1998-2011) in order to examine the appropriateness of the published factor structure of the DECA for this sample and to test for measurement invariance between the English- and Spanish- language versions. Since the identification of children

with behavioral concerns was the focus of the project, the series of confirmatory factor analyses (CFA; Bollen, 1989) were only conducted for the Behavioral Concerns subscale of the DECA (the individual items are listed in Table 1). For all CFA analyses: the data were specified as categorical and a robust weighted least squares (WLS) estimator was employed, WLSMV, with mean- and variance-adjusted $\chi^2$ (Muthén, 1993); and the *Mplus* syntax TYPE= COMPLEX was used to account for children being nested within classrooms (Muthén & Muthén, 1998-2010).

For all CFA analyses, approximate fit indices were used to assess the fit of the overall model to the data. Specifically, the Comparative Fit Index (CFI > .95; Bentler, 1990), an incremental fit index, and the root mean square error of approximation (RMSEA < .06; Steiger, 1990), a parsimony corrected index, of each model were evaluated. Given the categorical nature of the data and the robust WLS estimators that were employed, the weighted root mean square residual (WRMR < 1; Yu & Muthén, 2002) was also considered. If the other fit indices were adequate, a model with a significant chi-square test of model fit (p < .05) was considered acceptable since the $\chi^2$ statistic has been found to be unreliable with large samples (Bollen & Long, 1993). Additionally, theoretical justifications grounded in the extant literature were also considered when accepting or rejecting models.

**Research question 1: Is the published factor structure for the Behavioral Concerns subscale of the DECA appropriate for a sample of diverse, low-income preschool children?**

A CFA approach was used to examine the fit of the published factor structure to the English-language group. The published factor structure of the DECA Behavioral

Concerns subscale includes the ten categorical indicators (individual items on the DECA) that are explained by one continuous latent factor (Behavioral Concern). This CFA model is represented in Figure 1. Based on the results from the initial CFA, a series of exploratory factor analyses were conducted on the English language group in order to derive a factor structure appropriate for this low-income, diverse sample.

The most parsimonious factor solution, derived through exploratory factor analysis, was chosen because it met the following criteria: it (a) approximated simple structure as reflected in maximum hyperplane count (Gorsuch, 1983), (b) satisfied the constraints for the number of factors for retention [as estimated by minimum-average partialing (Velicer, 1976) and scree test (Cattell, 1966)], (c) retained salient loading items for each factor (loadings > .40; Gorsuch,1983), and (d) made theoretical sense in terms of the extant literature. The final factor structure derived via exploratory factor analysis was then examined by means of a confirmatory factor analysis.

**Research Question 2: Is the factor structure of the Behavioral Concerns subscale invariant across English- and Spanish-language forms of the DECA?**

To test for the most basic form of invariance, configural invariance (whether the same factor structure is manifested equivalently in each group), the newly derived factor structure was simultaneously fitted to the covariance structure of the English- and Spanish-language groups. Using a multiple group confirmatory factor analysis approach, the same (newly derived) factor structure was applied to both language groups while allowing all parameters to be estimated freely (Horn & McArdle, 1992).

To test for metric invariance (whether the latent construct explains the indicators in the same way across groups) the same, newly derived factor structure was applied to

both language groups and all of the unstandardized factor loadings and thresholds were

constrained equal across the two language groups, using a multiple group comparison

approach (Horn & McArdle, 1992). When data are categorical, tests of metric invariance

involve constraining the unstandardized factor loadings for each group being compared to

equality, to test for "weak" invariance and constraining the thresholds for each group

being compared to equality, to test for "strong" invariance (Meredith, 1993). While the

parameters, such as loadings and thresholds, can be released and constrained separately

when data are continuous (due to the independence of means and variances), the two

must be considered together when data are categorical, because the means and variances

of categorical data are not independent (Muthen & Muthen, 1998-2010).

**Chapter 3: Results**

The ten Behavioral Concerns items were examined for skewness and kurtosis for the English- and Spanish-language groups separately. First, the frequency distribution of item endorsement was examined for each language group and histograms were created for each item individually (separated by language). Visual inspection revealed that most items, with the exception of 23, 27 and 35, were positively skewed. Given the unequal groups (English = 5197 & Spanish = 860) it was best to compare each item's percentage of endorsements for each response category (Table 2). Additionally, visual inspection of the histograms (separated by language) suggested that none of the percentages were extremely different for any of the items. A skew index (SI) and kurtosis index (KI) was calculated for each item and problematic items were identified (SI >2 & KI > 7; Chou & Bentler, 1995). The same three items appeared to be most problematic in terms of both their SI and KI in the English- and Spanish-language groups (Item 9: touch children/adults inappropriately; Item 15: use obscene gestures or offensive language; and Item 18: destroy or damage property). SI's for Items 9, 15 and 18 were 4.0, 3.3 and 2.2, respectively for the English-language group and 10.7, 6.0 and 3.5, respectively for the Spanish-language group. KI's for Items 9, 15 and 18 were 16.7, 11.6 and 4.6, respectively for the English-language group and 1.07, 38.2 and 14.4, respectively for the Spanish-language group. While the other items did not meet the criteria for being problematically skewed or kurtotic, many of their distributions were not normal ("*never*" was most frequently endorsed for the majority of  items, specifically Items 8, 9, 11, 15, 18 & 26; see Table 2).  Due to the violation of the normality assumption, the data were treated as categorical (Bollen, 1989).

Visual inspection also revealed that there were no endorsements of the most extreme response category ("*very frequently*") for Item 15 (use obscene gestures or offensive language) in the Spanish-language group, while there were 25 endorsements in the English-language group. When the dependent variable is categorical, the statistical software that was utilized for the multiple group confirmatory analyses, M*plus*, requires that each group have same number of response category options (Muthén & Muthén, 1998-2010). Since there were zero endorsements for the fifth category for Item 15, the Spanish-language group would appear to only have four response category options. Therefore, it was necessary to collapse the category "*very frequently*" with the "*frequently*" category in the English-language group for Item 15 so that it would have the same number of response categories as the Spanish-language group.

**Research Question 1: Is the published factor structure for the Behavioral Concerns subscale of the DECA appropriate for a sample of diverse, low-income preschool children?**

When the published factor structure was applied to the English-language group (*n* = 5,197), the model resulted in poor fit, $\chi^2$ (88) = 4546.354, p < 0.0001, CFI = 0.709, TLI = 0.0801, RMSEA = 0.129, WRMR = 4.325. Despite the inadequate fit of the model, the internal consistency of the subscale was acceptable (Cronbach alpha = .74) and all of the items significantly loaded on the latent construct, Behavioral Concerns.

**Exploratory factor analysis for the English-language form.** Since the published factor structure of the Behavioral Concerns subscale did not fit these data adequately, a series of exploratory factor analyses were performed using the items from the English-language form to determine whether there was a more appropriate underlying

factor structure for the measure in this sample. To do this, the English-language sample was randomly split into two mutually exclusive subsamples [an index sample used for exploratory factor analyses (*n*= 2,598) and a reserve sample used for confirmatory factor analyses (*n*= 2,599)]. Prior to subjecting the index sample to exploratory factor analyses, MicroFACT 3.0 (Waller, 2001) was used to calculate polychoric item correlations (Olsson, 1979) and to smooth the matrix for nonsingularity and positive semidefiniteness.

A series of exploratory factor analyses were performed with the index sample (*n*= 2,598) to examine the underlying factor structure of the English-language form. In the publisher's development of the DECA, all 37-items were never subjected to an exploratory factor analysis simultaneously. Therefore, in the present study a set of common factor analyses were performed, separately for the Protective Concerns subscales and for the ten items from the Behavioral Concerns subscale (of interest to the present study was the Behavioral Concerns subscales, so further findings related to the Protective Concerns subscales are not presented here). For the ten Behavioral Concerns items, exploratory factor analysis suggested that a two-factor varimax solution resulted in the most parsimonious solution. Factors and component items are presented in Table 3. Factor I was named Externalizing Behavioral Concerns because it consisted of eight items related to externalizing behavior problems (e.g., has temper tantrums). Factor II, Internalizing Behavior Concerns, consisted of two items representing internalizing behaviors (has no reaction to children/adults & fails to show joy or gladness at happy occasion). Internal consistency for the externalizing factor was acceptable (Cronbach alpha = .81), but the internal consistency for the internalizing factor was low (Cronbach alpha = .63).

**Confirmatory factor analysis for the English-language form**. To confirm the newly derived structure for Behavioral Concerns, a confirmatory factor analysis in M*plus* was performed with the 10 Behavioral Concerns items from the English-language only reserve sample (*n*= 2,599). While the fit indices suggested that the model fit was not quite adequate, $\chi^2$ (34) = 878.074, p < 0.0001, CFI = 0.882, TLI = 0.843, RMSEA = 0.098, WRMR = 2.601, all of the fit indices suggested slight improvement when compared to the fit indices of the one-factor solution (increases in the CFI and TLI as well as decreases in the RMSEA and WRMR). However, statistical difference tests could not be performed because the two-factor and one-factor models were not nested (Muthen & Muthen, 1998-2010). Additionally, there is support in the literature for the distinction between internalizing and externalizing behavior problems (i.e. Bulotsky-Shearer, Fernandez, Dominguez, & Rouse, 2011; Campbell 2002; Shaw, Vondra, Dowdell-Hommerding, Keenan, & Dunn, 1994).Therefore, due to the improvement in the fit indices and theoretical justifications in the extant literature, the two-factor solution was retained.

**Research Question 2: Is the factor structure of the Behavioral Concerns subscale invariant across English- and Spanish-language forms of the DECA?**

**Configural invariance between the English- and Spanish-language forms**. Using the newly derived two-factor solution for the Behavioral Concerns subscale, two tests of invariance were performed: configural and metric. First, to test for configural invariance between the English- (*n* = 5,197) and Spanish-language (*n*= 860) forms, the same two-factor structure was simultaneously fitted to the covariance structure of the two language groups (using a multiple group comparison approach), but no other constraints

were placed on the models (all parameters were freely estimated). This model resulted in adequate fit, $\chi^2$ (105) = 1170.471, p < 0.0001, CFI = 0. 907, TLI = 0.920, RMSEA = 0.058, WRMR = 3.940, suggesting that configural invariance was plausible between the two language forms. Therefore, it could be concluded that the two-factor structure was manifested equivalently in the English- and Spanish-language forms of the DECA Behavioral Concerns subscale.

**Metric invariance between the English- and Spanish-language forms**. Since the initial form of invariance, configural invariance was achieved, it was appropriate to test for a stronger form for invariance, metric invariance. In this model the unstandardized factor loadings and thresholds were constrained equal across the entire sample of the two language groups, using a multiple group comparison approach. This constrained model resulted in adequate fit, $\chi^2$ (103) = 1138.679, p < 0.0001, CFI = 0.910, TLI = 0.921, RMSEA = 0.058, WRMR = 3.918. Since the most parsimonious model, with all of the factor loadings and thresholds constrained equal between groups, resulted in adequate fit, it was not necessary to release any constraints or explore less parsimonious models (Meredith, 1993). Therefore, it could be concluded that the constructs, Externalizing and Internalizing Behavioral Concerns, manifested themselves in the same way for both the English- and Spanish-language forms of the DECA.

**Chapter 4: Discussion**

The present study was the first to rigorously examine the measurement properties of the Behavioral Concern subscale of the DECA, to evaluate the appropriateness of its use within a large sample of diverse, low-income preschool children, and to empirically test for invariance between the English- and Spanish-language forms. The published factor structure could not be confirmed which provides evidence that the published factor structure of the subscale did not adequately represent the data for this low income, diverse sample. Consequently, a new factor structure was derived by means of exploratory factor analysis, resulting in a two-factor solution (Externalizing and Internalizing Behavioral Concerns). Test of configural and metric invariance revealed that with this two-factor structure, the English- and Spanish-language forms of the DECA Behavioral Concerns subscale reflected the construct, Behavioral Concerns, in the same way.

**Appropriateness of Published Factor Structure**

**Confirmatory factor analysis for the English-language form.** Consistent with the hypotheses, the published factor structure could not be confirmed. In other words, the individual items did not accurately reflect the desired construct (Behavioral Concerns). This finding could be due to several reasons. First, the Behavioral Concerns subscale was not empirically derived originally by the publisher, but rather the ten subscale items were selected based on "their psychometric properties and their representation of a wide range of challenging behaviors" (LeBuffe & Naglieri, 1999). Additionally, there is limited empirical evidence to support the one-factor structure of the Behavioral Concerns subscale, because no previous study to date has examined the subscale empirically; all

previous factor analytic studies of the DECA have examined only the Protective Factor and its subscales and have excluded the Behavior Concerns subscale. While the published structure of the ten item Behavioral Concerns subscale did not fit the data from our sample, it is also quite possible since the published factor structure was not empirically derived in the standardization sample, that the structure might not appropriately fit the standardization sample either and this is a direction for future research.

Furthermore, this finding comports with prior research, which has revealed the difficulty of confirming a measure's published factor structure when the measure was not developed or standardized for use with low-income populations (Fantuzzo, Hightower, Grim, Montes, 2002; Fantuzzo, Manz, McDermott, 1998; Konold, Hamre & Pianta 2003; LeBoeuf, Fantuzzo, & Lopez, 2010). The demographic characteristics of the DECA's norming sample and the demographic characteristics of present study's sample are quite distinct. The present study sample consisted of low-income (100% met the federal guidelines for poverty), predominantly African American and Hispanic children (55.4% and 43.8%, respectively). In the DECA standardization sample only 25% of the children were low-income and the majority of the norming sample was White (76%).

Additionally, studies have revealed that teacher and parent ratings of children's behaviors problems vary by child race and income; specifically, minorities from low-income families tend to be rated higher on behavior problem scales by teachers and parents (Achenbach & Rescorla, 2000; Gross et al., 2006). The differences in the ethnic and racial backgrounds of the children in the two samples (the standardization sample

versus the sample of the current study) may potentially be related to differences in the prevalence or expression of behavior problems in the two different samples.

**Two-Factor Solution**

**Exploratory and confirmatory factor analysis of the English language-form.** A series of exploratory factor analyses revealed that the items did not form a unidimensional construct, as suggested by the published factor structure. Rather the EFA revealed a two-factor varimax solution. These two factors, Externalizing and Internalizing Behavior Concerns comport with the extant literature on problem behaviors, which makes a distinction between externalizing and internalizing behavior problems. In the literature, externalizing behavior problems typically pertain to outward acts of aggression, disruption, tantrums, and over activity and internalizing behavior problems are typically defined by shyness, flat affect, and social withdrawal (e.g., Achenbach & Edelbrock, 1981; Campbell 2002; Hinshaw, Han, Erhardt & Huber, 1992). When the published, one-factor structure was applied, the structure could not be confirmed. However, when the newly derived, two-factor solution was applied the model resulted in fit statistics that were better than those generated by the published factor structure. Therefore, it could be concluded that this two-factor structure better represented the data.

**Configural and metric invariance.** The newly derived, two-factor solution was invariant between the English- and Spanish-language forms of the DECA. Therefore it can be concluded that the constructs, Externalizing and Internalizing Behavioral Concerns, manifested themselves in the same way across both language forms.

**Limitations and Future Directions**

While the present study contributed to the literature on preschool problem behavior and measurement invariance by using an entire cohort of culturally and linguistically diverse children from a large, urban Head Start program, limitations of the present study must be acknowledged. Because this is the first study to empirically examine the psychometric properties of the Behavioral Concern subscale and since the current sample came from one Head Start program, the generalizability of the findings may be limited to populations of predominantly African-American and Hispanic children from low-income families living in urban areas. Therefore, further examination of the fit of the published factor structure in other early childhood populations is warranted.

Additionally, the internal consistency for the Internalizing Behavioral Concerns factor was low. This is likely because the factor was only comprised of two items. However, it is important to acknowledge that examination of the underlying factor structure of the Behavior Concerns subscale in the present study was exploratory. Therefore, this newly derived factor structure should not be used for educational, research or clinical purposes.

**Conclusions and Implications**

The present study contributes to the literature on culturally and linguistically appropriate measures of preschool children, especially those from low-income and diverse backgrounds. Furthermore, findings regarding the fit of the published factor structure of the DECA's Behavioral Concerns and the psychometric properties of the newly derived two-factor solution can be used to inform practice. In the present study, the published one-factor structure of the Behavioral Concerns subscale did not fit the data for

a diverse, low-income sample. Due to this lack of fit, tests to examine the equivalence between the language forms, using the published factor structure, could not be performed. Therefore, comparisons should not be made between the language forms for scores obtained from the published factor structure because children's scores on the Spanish measures are obtained using the English form's norms.

In municipal Head Start programs with large Spanish-speaking populations of teachers, such as Miami-Dade, the measure is filled out electronically in either English or Spanish (depending on teachers' language preference). However, there is no indication or record of which language form the teachers used and common set of norms are used to derive children's scores, regardless of the language form. Therefore, is important for the publishers to provide a space where raters can record which language form is used so that distinctions can be made.

Additionally, this is the first study to rigorously evaluate the psychometric properties of the Behavioral Concerns subscale of the DECA. Interestingly, a different factor solution emerged when empirical means were employed. While the derivation of the new factor structure was exploratory and should not be used to inform referrals or interventions, the two-factor solution better comports with theory related to problem behaviors than does the one-factor published factor structure.

Findings revealed that the eight items that make up the newly derived Eternalizing Behavior Concern subscale generate reliable scores and that the subscale is invariant across the two language forms. Therefore, scores from these items can be compared between the English- and Spanish-language forms. However, the Internalizing Behavioral Concern (made up of only two items) subscale does not generate reliable scores.

Therefore, findings suggest as a recommendation, that the DECA could be used to identify children that display externalizing behaviors, but should not be used for identifying children with internalizing behavior problems. In general, national data suggests that there is an underreporting of children with emotional behavior problems by preschool teachers (Yoshikawa & Knitzer, 1997, Redden et. al., 1999), but this is especially true for children with internalizing behavior problems. Preschoolers that exhibit internalizing behavior problem are systematically missed by programs (Fantuzzo et. al., 2003). This is a critical problem because current research highlights a stronger relationship between internalizing behavior problems and poor social and academic school readiness outcomes than externalizing behavior problems (Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003). Therefore, the present study reveals a major limitation of the DECA: the Behavioral Concerns subscale does not provide sufficient information that can be used to screen and identify children who exhibit internalizing problems who are especially at risk for poor school readiness outcomes.

Rather than conducting future studies to explore the predictive and discriminant validity of the scores generated by the newly derived Externalizing and Internalizing Behavioral Concerns subscales, a more comprehensive measure of children's classroom behavior should be used in early childhood programs. In order to comport with theory and the extant literature, measures of classroom behavior should be multidimensional, contextually relevant, psychometrically sound and appropriate for use within diverse populations (Fantuzzo, Manz, McDermott, 1998; Konold, Hamre & Pianta 2003). One such measure is the Adjustment Scale for Preschool Intervention (ASPI; Lutz, Fantuzzo, McDermott, 2002). The ASPI is a contextually relevant measure of classroom behavior

problems which demonstrates reliable and valid scores in low-income samples and assesses a broad range of externalizing and internalizing behavior problems.

In summary, findings from the present study contribute to the literature in several ways. First, findings highlight the importance of empirically deriving and evaluating the psychometric properties of multidimensional measures, as well as ensuring that such measures are grounded in developmental theory, especially if measures are to be used as tools for early identification and intervention within early childhood programs. Additionally, the lack of fit (when the published factor structure was applied) confirms the importance of adhering to best practice in cross-cultural measurement development and empirically examining the latent constructs prior to assuming equivalence between language forms. Because of limitations of the DECA's Behavioral Concerns subscale revealed here, caution by early childhood programs should be taken when interpreting children's scores on this subscale. Rigorous empirical research is needed especially if measures are to be used for early identification and intervention practices within early childhood programs, so that children's mental health needs are accurately and appropriately identified.

# References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Achenbach, T. M. & Edelbrock, C. (1981). Behavioral problems and competencies reported byparents of normal and disturbed children age four through sixteen. *Monographs of the Society for Research in Child Development,* 46 (Serial No. 188).

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Administration for Children and Families. *Study Design and Data Tables for FACES 2006 Baseline Report*, Washington, DC: U.S. Department of Health and Human Services, 2008.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.

Barbarin, O. A. (2007). Mental health screening of preschool children: Validity and reliability of ABLE. *American Journal of Orthopsychiatry*, 77(3), 402-418.

Bentler, P. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bollen, K.A. (1989). *Structural equations with latent variables*. Wiley & Sons, New, NY.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage Publications, Inc.

Brinkman, T. M., Wigent, C. A., Tomac, R. A., Pham, A. V., & Carlson, J. S. (2007). Using the Devereux Early Childhood Assessment to identify behavioral risk and protective factors within a Head Start population. *Canadian Journal of School Psychology, 22*, 136-151.

Bulotsky-Shearer, R. J., Fernandez, V., Dominguez, X., & Rouse, H. L. (2011). Behavior problems in learning activities and social interactions in Head Start classrooms and early reading, mathematics, and approaches to learning. *School Psychology Review*, *40*(1), 39-56.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37(2), 231-242.

Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues (2nd ed.)*. New York, NY US: Guilford Press.

Cattell, R. B. (1966). *Handbook of multivariate experimental psychology*. Chicago, IL: Rand McNally.

Chou, C.-P., & Bentler, P. M. (1995). Estimation and tests in structural equation modeling. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 37–55). Thousand Oaks, CA: Sage.

Cooper, J. L., Aratani, Y., Knitzer, J., Douglas-Hall, A., Masi, R., Banghart, P. L., & Dababnah, S. (2008). Unclaimed children revisited: The status of children's mental health policy in the United States.

Copple, C., & Bredekamp, S. (2009). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. National Association for the Education of Young Children. 1313 L Street NW Suite 500, Washington, DC 22205-4101.

Denham, S. A. (2006). Social-Emotional Competence as Support for School Readiness: What is it and how do we assess it?. *Early Education & Development, 17*(1), 57-89.

Deyo, R. A. (1984). Pitfalls in measuring the health status of Mexican-Americans: Comparative validity of the English and Spanish Impact Profile: *American Journal of Public Health, 74*, 569-573.

Drasgow, F. (1984). Scrutinizing psychological tests: measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin, 95,* 135-135.

Fantuzzo, J., Bulotsky, R. J., McDermott, P., Mosca, S., & Lutz, M. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32*, 185-203.

Fantuzzo, J., Hightower, D., Grim, S., & Montes, G. (2002). Generalization of the Child Observation Record: A validity study for diverse samples of urban, low-income preschool children. *Early Childhood Research Quarterly, 17*(1), 106-125.

Fantuzzo, J., Manz, P. H., & McDermott, P. (1998). Preschool version of the social skills rating system: An empirical analysis of its use with low-income children. *Journal of School Psychology, 36*, 199-214.

Feil, E. G., Small, J. W., Forness, S. R., Serna, L. A., Kaiser, A. P., Hancock, T. B., & ... Lopez, M. L. (2005). Using different measures, informants, and clinical cut-off points to estimate prevalence of emotional or behavioral disorders in preschoolers: Effects on age, gender, and ethnicity. *Behavioral Disorders*, 30(4), 375-391.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J. M., Richardson, R., & Sivan, A. (2006). The equivalence of the child behavior Checklist/1 1/2-5 across parent race/ethnicity, income level, and language. *Psychological Assessment, 18*, 313-323.

Hinshaw, S. P., Han, S. S., Erhardt, D. & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child Psychology*, 21, 143-150.

Horn, J. L. (1991). ''Comments on 'Issues in Factorial In-variance," in *Best Methods for the Analysis of Change,* Linda M. Collins and John L. Horn, Washington, DC: American Psychological Association, 114–125.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide in measurement invariance in aging research. Experimental Aging Research, 18, 117−144.

Jaberg, P. E., Dixon, D. J., & Weis, G. M. (2009). Replication evidence in support of the psychometric properties of the Devereux Early Childhood Assessment. *Canadian Journal of School Psychology*, 24(2), 158-166.Kline, R. B. (2011) *Principles and Practice of Structural Equation Modeling.* The Guilford Press

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika,36*: 409-426.

Konold, T. R., Hamre, B. K., & Pianta, R. C. (2003). Measuring problem behaviors in young children. *Behavioral Disorders, 28,* 111-123.

Lavigne, J. V., Gibbons, R. D., Christoffel, K. K., Arend, R., Rosenbaum, D., Binns, H.,…Isaacs, C. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 204-214.

LeBoeuf, W. A., Fantuzzo, J. W., & Lopez, M. L. (2010). Measurement and population miss-fits: A case study on the importance of using appropriate measures to evaluate early childhood interventions. *Applied Developmental Science*, 14(1), 45-53.

LeBuffe, P. A., & Naglieri, J., A. (1999). *Devereux Early Childhood Assessment (DECA).* Lewisville, NC: Kaplan Press.

LeBuffe, P. A., & Naglieri, J., A. (2001). *Devereux Early Childhood Assessment – Spanish version (DECA-Sp).* Lewisville, NC: Kaplan Press.

Lien, M. T., & Carlson, J. S. (2009). Psychometric properties of the Devereux Early Childhood Assessment in a Head Start sample. *Journal of Psychoeducational Assessment*, 27(5), 386-396.

Lutz, M., Fantuzzo, J., & McDermott, P. (2002). Multidimensional assessment of emotional and behavioral adjustment problems of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly*, *17*, 338-355.

Meade, A.W., & Lautenschlager, G.J. (2004). A monte-carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11,* 60-72.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.

Muthén, B.O. (1993). Goodness of fit with categorical and other non-normal variables. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 205-243). Newbury Park, CA: Sage.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

National Association for the Education of Young Children (NAEYC). (2005). *Screening and assessment of young English-language learners: Supplement to the NAEYC and NAECS/SDE joint position statement on early childhood curriculum, assessment, and program evaluation.* Washington, DC:   Author.

National Association for the Education of Young Children. (2003) E*arly childhood curriculum, assessment and program evaluation.* Position statement. Washington, DC: Author.

National Center for Educational Statistics. (2004). *The condition of education.* Washington, DC: U.S. Department of Education.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443-460.

Pena, E. D.(2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development,78,*1255-1264.

Qi, C. H., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: A review of the literature. *Topics in Early Childhood Special Education, 23,* 188–216.

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report, 16*, 3–6.

Reynolds, A. J., & Ou, S. (2003). Promoting resilience through early childhood intervention. In S. S. Luthar, S. S. Luthar (Eds.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. 436-459). New York, NY US: Cambridge University Press.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: review of practice and implications. Human Resource Management Review, 18, 210–222.

Shaw, D. S., Vondra,J. I., Dowdell-Hommerding, K, Keenan, K & Dunn, M. (1994). Chronic family adversity and early child behavior problems: a longitudinal study of low income families./ourraa/*of Child Psychology and Psychiatry,* 35, 1109-1122.

Sinclair, E. (1993). Early identification of preschoolers with special needs in Head Start. *Topics in Early Childhood Special Education*, 13(2), 184-201.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102–111

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

U. S. Department of Health and Human Services. (2007). *Improving Head Start for School Readiness Act of 2007*. H.R. 1429: Washington, DC: U.S. Government Printing Office. Retrieved from http://www.govtrack.us/congress/billtext.xpd?bill=h110-1429.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 2, 4–69.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers'changing commitment to the organization during the first 6 months of work. *Journal of Applied Psychology, 78,* 557-568.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321–327.

Webster-Stratton, C., & Hammond, M. (1998). Conduct problems and level of social competence in Head Start children: Prevalence, pervasiveness, and associated risk factors. *Clinical Child and Family Psychology Review*, 1(2), 101-124.

Yu, C. Y., & Muth´en, B. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes* (Technical Report). Los Angeles: University of California at Los Angeles, Graduate School of Education & Information Studies.

Table 1

*Behavioral Concerns items on the DECA: English and Spanish*

**Q8**　**fail to show joy or gladness at a happy occasion?**

*dejó de expresar alegría o satisfacción en una ocasión alegre?*

**Q9**　**touch children/adults inappropriately?**

*tocó a otros niños o adultos de manera inapropiada?*

**Q11**　**have temper tantrums?**

*hizo una rabieta?*

**Q14**　**have no reaction to children/adults?**

*se mostró indiferente a otros niños o adultos?*

**Q15**　**use obscene gestures or offensive language?**

*utilizó gestos obscenos o lenguaje ofensivo?*

**Q18**　**destroy or damage property?**

*destruyó o dañó algún objeto?*

**Q23**　**have a short attention span (difficulty concentrating)?**

*su atención fue reducida (es decir, tuvo dificultad para concentrarse)?*

**Q26**　**fight with other children?**

*se peleó con otros niños?*

**Q27**　**become upset or cry easily?**

*se molestó o lloró con facilidad?*

**Q35**　**get easily distracted?**

*se distrajo con facilidad?*

Table 2

*Frequency of items endorsed across language form: Percent endorsed English (Spanish in parentheses)*

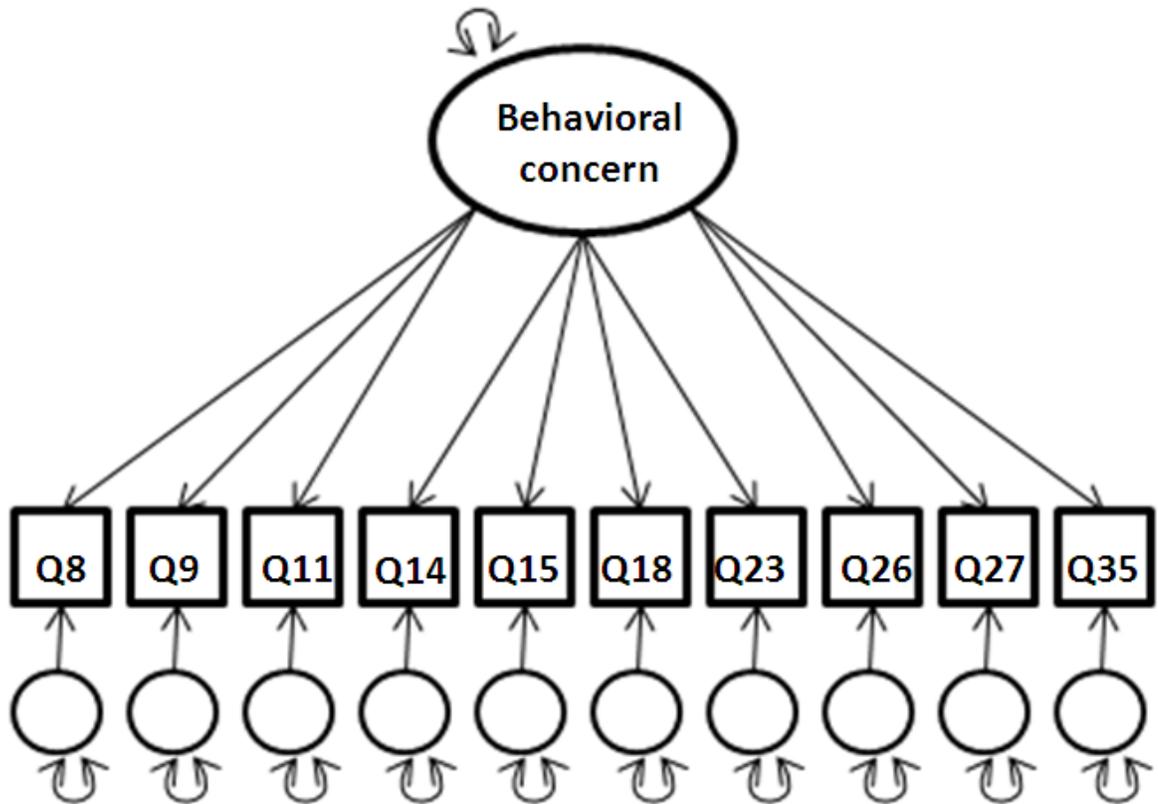| Item | 0 Never | 1 Rarely | 2 Occasionally | 3 Frequently | 4 Very Frequently |
|---|---|---|---|---|---|
| *Externalizing Behavioral Concerns* | | | | | |
| Q9   Touch inappropriately | 89.3 (98.4) | 5.9  (0.8) | 2.4 (0.2) | 1.5 (0.2) | 0.9 (0.3) |
| Q11  Temper tantrums | 57.6 (26.5) | 24.1 (26.5) | 12.7 (8.3) | 3.8 (3.0) | 1.8 (1.2) |
| Q15  Obscene gestures | 86.6 (95.8) | 7.4  (2.4) | 3.8 (1.4) | 2.2 (0.3) | N/A |
| Q18  Destroy or damage property | 73.3 (87.2) | 17.0 (9.1) | 6.2 (3.0) | 2.8 (0.5) | 0.8 (0.2) |
| Q23  Short attention span | 14.0 (16.6) | 35.2 (40.5) | 35.1 (33.1) | 12.1 (6.9) | 3.6 (2.9) |
| Q26  Fight with children | 73.3 (87.2) | 17.0 (9.1) | 6.2 (3.0) | 2.8 (0.5) | 0.8 (0.2) |
| Q27  Upset or cry easily | 14.0 (16.6) | 35.2 (40.5) | 35.1 (33.1) | 12.1 (6.9) | 3.6 (2.9) |
| Q35  Easily distracted | 41.9 (41.5) | 34.4 (40.3) | 17.0 (14.0) | 4.8 (3.4) | 1.9 (0.8) |
| *Internalizing Behavioral Concerns* | | | | | |
| Q8   Fail to show joy | 46.2 (49.5) | 33.0 (33.4) | 12.5 (2.4) | 6.0 (1.9) | 2.3 (0.7) |
| Q14  No reaction to children/adults | 46.0 (50.9) | 32.6 (34.8) | 14.9 (11.0) | 5.2 (2.8) | 1.3 (0.5) |

*Note.* When the dependent variable is categorical, *Mplus*, requires that each group have same number of response category options (Muthén & Muthén, 1998-2010). Since there were zero endorsements for the fifth category for Item 15, the Spanish-language group would appear to only have four response category options. It was necessary to collapse the category "*very frequently*" with the "*frequently*" category in the English-language group for Item 15 so that it would have the same number of response categories as the Spanish-language group.

Table 3

*Exploratory Factor Structure: Two-factor Solution*
*Newly Derived DECA Exploratory Factor Structure (Behavioral Concerns Subscales)*

| | | Behavioral Concerns | |
| --- | --- | --- | --- |
| Item | Label | Externalizing Behavior | Internalizing Behavior |
| 18 | Destroy or damage property | 0.78 | |
| 26 | Fight with other children | 0.73 | |
| 15 | Use obscene gestures or offensive language | 0.71 | |
| 11 | Have temper tantrums | 0.67 | |
| 9 | Touch children/adults inappropriately | 0.67 | |
| 27 | Become upset or cry easily | 0.56 | |
| 23 | Have a short attention span (difficulty concentrating) | 0.55 | |
| 35 | Get easily distracted | 0.53 | |
| 14 | Have no reaction to children/adults | . | 0.82 |
| 8 | Fail to show joy or gladness at a happy occasion | . | 0.68 |

*Note.* Final exploratory factor structure was a 2-factor, varimax orthogonal solution ($n = 2,598$ for the index sample).

*Figure 1.* Measurement model: Behavioral Concerns subscale of the DECA (published factor structure)