Early Childhood Research Quarterly 36 (2016) 366–378

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Using bifactor models to measure teacher–child interaction quality in early childhood: Evidence from the Caregiver Interaction Scale

Annemarie H. Hindman[a,*], Laura L. Pendergast[a], Rachel A. Gooze[b]

[a] Temple University, United States
[b] Child Trends, United States

ARTICLE INFO

Article history:
Received 1 August 2014
Received in revised form 18 January 2016
Accepted 19 January 2016
Available online 6 February 2016

Keywords:
Head Start
Teacher–child interaction
Factor analysis
Validity

ABSTRACT

Bifactor models have great promise to support the measurement of adult–child interaction in early childhood settings but are not frequently used in the field. This study explored whether a bifactor model fit teacher–child interaction data gathered from the Caregiver Interaction Scale (CIS; Arnett, 1989) in four cohorts of the recent Head Start Family and Child Experiences Survey (FACES) study (1997–2006). Analyses also examined concurrent validity of this approach using several teacher- and child-level variables. In total, 1422 Head Start classrooms were observed with the CIS. Factor analyses found that a bifactor model, featuring one factor for overall positive teacher–child interaction as well as two methodological factors accounting for whether items targeted appropriate or (reverse-coded) inappropriate behaviors, fit the data well, consistent with other recent work. Further, evidence of concurrent validity for this bifactor model of teacher-child interaction emerged with lead teachers' background factors (experience and CDA credential) and their global classroom quality, as well as children's prosocial skills. Overall, results illustrate both the utility and logistics of the bifactor model approach to measuring interaction quality in early childhood settings.

1. Introduction

A primary feature of a high-quality early learning experience is a positive affective relationship between teachers and children (Maldonado-Carreño & Votruba-Drzal, 2011; Pianta, 1999). Specifically, ecological and attachment theories (Bornstein & Tamis-LeMonda, 1989) both suggest that children are more likely to thrive, cognitively and emotionally, when their caregivers are (a) positive and enthusiastic about them; (b) regularly engaged with them; (c) aware of and sensitive to their particular intellectual and emotional needs; (d) systematic in the creation, explanation, and application of rules and guidelines; and (e) skilled in helping them begin to regulate their own emotions and behavior (Ainsworth, Blehar, Waters, & Wall, 1978; Pianta & Steinberg, 1992). Such positive teacher–child relationships may be especially important for children in poverty. Ample evidence shows that these relationships can support vulnerable learners' positive social and emotional development (Bierman et al., 2014; Curby, Rimm-Kaufman, & Ponitz, 2009; Garner, Mahatmya, Moses, & Bolt, 2014) and reduce problem behaviors (Raver et al., 2009; Smith, Lewis, & Stormont, 2011; Webster-Stratton, Reid, & Stoolmiler, 2008). However, the affective quality of teacher-child interactions varies widely across preschools, particularly in low-income settings (Garner et al., 2014). Consequently, the field would benefit from further focus on accurately measuring and, ultimately, enhancing this facet of the quality of early care.

2. Measuring the affective quality of early care environments: bifactor models

Because teacher–child interaction is a complex phenomenon, accurately operationalizing and measuring it can present significant methodological hurdles. A key issue involves isolating which specific components, or dimensions, among the multiple, interconnected processes that comprise interactions actually merit measurement. Consider how teacher–child interactions unfold during a typical classroom activity: a preschool teacher must simultaneously structure the space and schedule of the class in ways that foster children's attention and motivation; complement children on appropriate behaviors; respond sensitively, firmly, and promptly to transgressions; help children manage their own emotions and navigate charged exchanges with peers; and engage

* Corresponding author at: Annemarie Hindman, Psychological Studies in Education, College of Education, Temple University, 1301 Cecil B. Moore Ave, Philadelphia, PA 19122, United States.
E-mail address: ahindman@temple.edu (A.H. Hindman).

http://dx.doi.org/10.1016/j.ecresq.2016.01.012
0885-2006/

children through meaningful instruction. Clearly, many processes and nuances are involved in teacher–child interactions, even over a brief span of time.

On one hand, the task of measuring the complex construct of teacher–child interaction may be thought of as capturing each of a collection of distinct but entangled dimensions or elements (Grusec & Davidov, 2010), such as positive interaction, punitiveness, and personal engagement with children (Arnett, 1989). Such efforts generally employ exploratory factor analysis (EFA; Comrey & Lee, 1992) to isolate distinct but correlated sets of items (i.e., factors) that each reflect a single specific underlying aspect of the complex construct. Subsequent validation models generally regress distal outcomes on all factors from the EFA at once, so that only the unique variation in each factor explains variability in the outcomes. However, this approach to the conceptualization of interaction-related data can be problematic. One issue is that proposed factors may be so conceptually and practically intertwined (e.g., teachers have to be engaged to interact positively or punitively) that they essentially reflect a single underlying factor (Chen, West, & Sousa, 2006). A second challenge is that one proposed factor (e.g., positive interaction) may represent the opposite end of a continuum from another factor (e.g., punitiveness) rather than a truly distinct dimension (Arnett). A third concern is that including multiple, correlated factors in the same regression model may leave little unique variance in each to explain outcomes and can consequently attenuate concurrent or predictive validity (Rudasill, Gallagher, & White, 2010).

On the other hand, a more plausible and holistic approach to conceptualizing the complexity of teacher-child interaction involves including an overarching factor that accounts for most of the variance in the data, along with multiple, independent sub-factors that account for some additional variance. For example, Hamre, Hatfield, Pianta, and Jamil (2014) have framed teacher-child interaction quality as measured by the CLASS (Pianta, LaParo, & Hamre, 2008) as largely driven by the overarching construct of teacher responsiveness, but they also include multiple sub-factors including consistent management and conceptual engagement that explain a smaller, unique portion of the variance as well. Overall, this approach cohesively captures, in a single model, the global and specialized dimensions of interaction.

While EFA does not allow for the exploration of such a scenario, bifactor modeling, a SEM technique, represents a highly effective tool for working with such multidimensional data (Reise, 2012). Essentially, at the item level, bifactor modeling allows researchers to simultaneously examine multiple latent factors that influence item-level variance (Canivez, in press). With this technique, each item is hypothesized to load onto one general factor (often tapping a conceptual construct), as well as one of two or more additional factors (often tapping nuanced conceptual or methodological features); all of these factors are modeled as uncorrelated with each other within the overall model (Gibbons & Hedeker, 1992). In this way, a bifactor model can tease apart conceptual (and potentially methodological) variance to better identify meaningful factors within a single, nuanced analysis.

A number of recent papers have successfully applied this technique to new early childhood measures (McDermott et al., 2011) or refined our understanding of those in use for many years (Betts, Pickart, & Heistad, 2011; Burke et al., 2014; Hamre et al., 2014; Lakin & Gambrell, 2012; Norwalk, DiPerna, & Lei, 2014; Park, Dimitrov, Das, & Gichuru, 2014; Varni, Beaujean, & Limbers, 2013). However, because this methodology is not yet widely used in the examination of early care interaction quality, the current paper offers an example of this approach. We used the Caregiver Interaction Scale (CIS; Arnett, 1989) as an illustration, given its familiarity in the field and its highly uncertain, contested factor structure.

3. The Caregiver Interaction Scale

The CIS includes 30 items describing positive or negative teacher behaviors (e.g., "The teacher encourages children to try new experiences" or "The teacher punishes children without an explanation," respectively). An observer rates the agreement between each statement and the teacher's behavior on a scale from 1 (not at all like the teacher) to 4 (very much like the teacher). The CIS has been widely used in smaller scale research projects across the U.S. (Bracken & Fischel, 2006; Conners-Burrow, Whiteside-Mansell, McKelvey, Virmani, & Sockwell, 2012, Conners-Burrow et al., 2013; Powell, Son, File, & San Juan, 2010; Rappolt-Schlichtmann et al., 2009), in statewide evaluations of early care (e.g., Wisconsin, Massachusetts, Missouri, and Pennsylvania; see Miller & Bogatova, 2009), and internationally (Sylva et al., 2006). Further, many publically available early childhood datasets use the CIS, including the Head Start Impact Study (HSIS; Puma et al., 2010), FACES (Bulotsky-Shearer, Wen, Faria, Hahs-Vaughn, & Korfmacher, 2012; Halle, Hair, Wandner, & Chien, 2012; Hindman, Cromley, Skibbe, & Miller, 2011; McWayne, Cheung, Wright, & Hahs-Vaughn, 2012; McWayne, Hahs-Vaughn, Cheung, & Wright, 2012; Son, Kwon, Jeon, & Hong, 2013), and the Early Childhood Longitudinal Study-Birth cohort (ECLS-B; Fram & Kim, 2012; Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013).

Despite the popularity of the CIS, research using this tool has struggled to coherently aggregate the items into reliable, replicable factors. The tool was first published by Arnett (1989); notably, this initial study was not focused on measure development. Arnett used principal components methods in a small sample (item-to-participant ratio was approximately 1:2) to derive four factors from 26 items: positive interaction, punitiveness, detachment, and permissiveness. Arnett did not report which items were assigned to each subscale or whether to combine the subscales to create an overall score (i.e., a second-order factor). Over time, few studies conducted additional, rigorous factor analyses to explore these open questions; instead, projects have made a wide range of choices for handling the data. For example, some (Powell et al., 2010) have used just one of the original factors, whereas others (Kontos, Howes, & Galinsky, 1996) have sorted items into fewer factors, and still others have used a total score (Hindman, Skibbe, Miller, & Zimmerman, 2010) or added four additional items related to support for child independence as an additional factor (Resnick & Zill, 2001). Not surprisingly, these varied approaches to the factor structure of the measure have yielded inconsistent results regarding its predictive validity (Curby, LoCasale-Crouch et al., 2009), with some studies finding modest predictive links to children's social and academic outcomes (Loeb, Fuller, Kagan, & Carrol, 2004; Peisner-Feinberg et al., 2001), while others uncover null associations (Hindman et al., 2010; Lisonbee, Mize, Payne, & Granger, 2008; Zill et al., 2003).

Also complicating its use, the CIS includes a methodological confound in that all items tapping positive interaction describe appropriate behaviors (e.g., "Teacher speaks warmly to children"), whereas all items on the other factors (i.e., punitive, permissive, detached) describe inappropriate teacher behaviors (e.g., "Teacher seems critical of children"). The Arnett (1989) study did not specify whether researchers should reverse-code the inappropriate-behavior (also referred to as negatively worded) items for analysis, although later work has generally done so. This recoding facilitates the combination of items across subscales, as higher scores on any item reflect more supportive practices, but it may oversimplify the distinction between appropriate and inappropriate practices. Specifically, recoding inappropriate-practice items and then combining them with appropriate-practice items implies that observers will rate the absence of negativity (e.g., rarely reprimands children harshly) similarly to the presence of positivity