Infant Behavior & Development 39 (2015) 21–34

Contents lists available at ScienceDirect

Infant Behavior and Development

The epidemiology of observed temperament: Factor structure and demographic group differences

Michael T. Willoughby[a,*], Cynthia A. Stifter[b], Nisha C. Gottfredson[c], the Family Life Project Investigators[1]

[a] Education & Workforce Development, RTI International, United States
[b] Human Development & Family Studies, Pennsylvania State University, United States
[c] Center for Developmental Science, University of North Carolina at Chapel Hill, United States

ARTICLE INFO

Article history:
Received 19 May 2014
Received in revised form
29 December 2014
Accepted 10 February 2015
Available online 27 February 2015

Keywords:
Temperament
Reactivity
Regulation
Latent variable
Confirmatory factor analysis

ABSTRACT

This study investigated the factor structure of observational indicators of children's temperament that were collected across the first three years of life in the Family Life Project ($N = 1205$) sample. A four-factor model (activity level, fear, anger, regulation), which corresponded broadly to Rothbart's distinction between reactivity and regulation, provided an acceptable fit the observed data. Tests of measurement invariance demonstrated that a majority of the observational indicators exhibited comparable measurement properties for male vs. female, black vs. white, and poor vs. not-poor children, which improved the generalizability of these results. Unadjusted demographic group comparisons revealed small to moderate sized differences (Cohen ds = |.23 – .42|) in temperamental reactivity and moderate to large sized differences (Cohen ds = –.64 – –.97) in regulation. Collectively, demographic variables explained more of the variation in regulation ($R^2 = .25$) than in reactivity ($R^2 = .02$–$.06$). Follow-up analyses demonstrated that race differences were substantially diminished in magnitude and better accounted for by poverty. These results help to validate the distinction between temperamental reactivity and regulation using observational indicators.

1. Introduction

Numerous models and definitions of temperament exist (Goldsmith et al., 1987; Rothbart, Derryberry, & Posner, 1994; Strelau, 1994), a consensus definition characterizes temperament as individual differences in behavioral tendencies that are evident early in life and reflect early biological predispositions that are shaped by contextual experience (Rothbart & Bates, 1998). Scholars from a wide range of disciplines are empirically interested in the construct of temperament. For example, scientists who study prenatal development routinely consider temperament as an early outcome that is associated with early exposure to toxicants, (non)prescription drugs, and general stressors (Blair, Glynn, Sandman, & Davis, 2011; Grey, Davis, Sandman, & Glynn, 2013; Mayes, 2002; Richardson, Goldschrmidt, & Willford, 2008; Schuetze, Molnar, & Eiden, 2012; Weiss, Jonn-Seed, & Harris-Muchell, 2007). Cognitive neuroscientists have characterized temperament as "model area of study" for questions focused on the inter-relations of cognitive and emotional functioning (Bell & Wolfe, 2004; Henderson & Wachs,

* Corresponding author at: RTI International, Hobbs #349, Research Triangle Park, NC 27709-2194, United States. Tel.: +1 9195416622.
E-mail address: mwilloughby@rti.org (M.T. Willoughby).
[1] The members of the Family Life Project Investigators are listed in the Acknowledgements section.

http://dx.doi.org/10.1016/j.infbeh.2015.02.001
0163-6383/

2007; Wolfe & Bell, 2004). Clinical psychologists and psychiatrists are concerned with the prognostic value of temperament to forecast emergent psychopathology in early and middle childhood (Bijttebier & Roeyers, 2009; Martel, 2009; Muris & Ollendick, 2005; Nigg, 2006; Rapee & Coplan, 2010).

Parent-report inventories remain the most widely utilized method for measuring temperament (Garstein et al., 2012). The popularity of parent-reports of temperament is due both to their ease of administration and acknowledgment of parents' unique vantage of their children's behavioral tendencies across time and settings. Due to the ease of administration and scoring, large-scale studies have typically relied exclusively on parent report questionnaires to measure temperament (Henrichs et al., 2009; Kaplan et al., 2014; Sanson, Prior, Garino, Oberklaid, & Sewell, 1987; Wessman et al., 2012). Studies involving parent reported inventories of temperament have provided empirical support for sub-dividing the construct of temperament into two broad domains—reactivity and regulation (Putnam & Stifter, 2008). Whereas reactivity has been defined as "the speed, strength, and valence [positive or negative] of an individual's characteristic response to stimulation", regulation has been defined as "behaviors the individual uses to control their behavioral and emotional reactions to sources of both positive and negative stimulation" (Henderson & Wachs, 2007; p. 400).

Despite their ease of use, parent reports of temperament have been criticized because they tend to be weakly associated with observed temperamental behaviors (Forman et al., 2003; Seifer, Sameroff, Dickstein, Schiller, & Hayden, 2004; Vaughn, Bradley, Joffe, Seifer, & Barglow, 1987; Vaughn, Taraldson, Crichton, & Egeland, 1981). Characteristics of parents (personality, psychopathology, stress) and parent–child interaction quality have been identified as potential threats to the validity of parent reports of temperament (Forman et al., 2003; Leerkes & Crockenberg, 2003; Mebert, 1991; Parade & Leerkes, 2008; Sameroff, Seifer, & Elias, 1982). Conversely, the novelty and artificiality of observational tasks that are used to elicit temperamental behaviors, the variations in time scale in which behaviors are observed (seconds to minutes for observational measures vs. days/weeks for ratings), and contextual differences all represent alternative explanations for the weak associations between parent reported and observed temperament. Regardless of the specific reasons, the lack of a strong relation between parent report and observed temperament undermines conventional measurement wisdom, which implies that there is scientific value in aggregating information across informant and methods (Campbell & Fiske, 1959; Podsakoff, MacKenzie, & Podsakoff, 2012). Specifically, to the extent that parent reports and observed temperament are weakly associated, there exists little shared variation to define the latent constructs of interest.

Data collectors who conduct laboratory or in-home visits represent another potential source of information on children's temperament that are not subject to the same concerns related to parent rated temperament (e.g., Matheny, 1983). Data collector's global impressions of children's temperament complements information that is obtained from microsocial (e.g., second-by-second) coding of children's responses to challenge tasks (Carnicero, Perez-Lopez, Del Carmen, Salinas, & Martinez-Fuentes, 2000; Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011; Stifter & Corey, 2001). Indeed, we previously demonstrated the merits of using home visitors reports of temperament in the same sample as will be used in the current study (Stifter, Willoughby, Towe-Goodman, & Investigators, 2008). Those results demonstrated that while there was convergence across parents, home visitors, and observation measures of infant positivity, only home visitor and observational measures converged with respect to the measurement of infant negativity. Presumably, home visitor impressions were influenced by their administration of challenge tasks, while parent reports took into account other sources of information.

The first objective of this study was to test whether home visitors ratings and multiple observational (including performance-based) tasks might be used to build latent constructs of three subdomains of temperamental reactivity (fear, anger, activity level), as well as a single domain of regulation. This builds on a small number of recent studies that have begun to systematically investigate the structure and stability of temperament in early and middle childhood using exclusively observational measures (Durbin, Hayden, Klein, & Olino, 2007; Dyson, Olino, Durbin, Goldsmith, & Klein, 2012; Gagne et al., 2011; Kotelnikova, Olino, Mackrell, Jordan, & Hayden, 2013). Our proposed work most closely resembles that of Dyson et al. (2012) and Kotelnikova et al. (2013), both of whom utilized factor analytic methods to test the structure of observed temperament in early and middle childhood, respectively. While the current study differed in important ways from those studies—e.g., sample acquisition (representative vs. convenience), study design (longitudinal vs. cross-section), child age (birth – 3 years vs. early or middle childhood), nature and scope of coding systems, and the nature of constructs considered—the guiding premise of the current study was the same as those previous studies, namely to provide a vantage of the structure of temperament that was independent of parent reports.

A prevailing assumption in the literature is that observational tasks designed to elicit temperamental reactivity and regulation work equally well for all populations. Tests of measurement invariance provide one means of evaluating this assumption. Measurement invariance involves testing a sequence of models that impose increasingly stringent requirements regarding the measurement equivalence of a set of indicators of a latent construct across groups (Meredith, 1993). Here, we exploit the characteristics of the current sample (large *N*, over-sampling of low income and African American families) to test the measurement invariance of observational measures of temperamental reactivity and regulation as a function of child sex, race, and household poverty status. To be clear, testing whether observational indicators of temperament cohere in comparable ways across children of demographic groups provides one means of evaluating the generalizability of the organizing distinction between reactivity and regulation. We are unaware of any previous studies that have tested these questions using observational data, though this approach has been utilized with parent- and self-report measures (Carter, Briggs-Gowan, Jones, & Little, 2003; Kim, Brody, & Murry, 2003; Zimprich & Mascherek, 2012).

To the extent that temperamental constructs of reactivity and regulation exhibit at least partially strong invariance across groups, this ensures that any observed group differences are meaningful. A tertiary objective of this study was to evaluate