

Measurement Reactivity and Fatigue Effects in Daily Diary Research With Families

Bridget M. Reynolds, Theodore F. Robles, and Rena L. Repetti
University of California, Los Angeles

Methodological challenges associated with measurement reactivity and fatigue were addressed using diary data collected from mothers ($n = 47$), fathers ($n = 39$), and children ($n = 47$; 8–13 years) across 56 consecutive days. Demonstrating the feasibility of extended diary studies with families, on-time compliance rates were upward of 90% for all family members, with only minor within-person declines in weekday compliance over time. Multilevel models revealed slight decreases in mother and father daily reports of parent–child conflict and warmth across days, suggesting possible measurement reactivity. Global perceptions of parent–child involvement, measured via a 1-time survey at baseline, moderated change in parent, but not child, diary reports of conflict and warmth. Finally, weakening agreement between mother and child diary reports of conflict and strengthening of positive within-person associations between child-reported negative mood and same-diary ratings of parent–child conflict indicate potential fatigue-related declines in response accuracy. Although generally minimal, observed measurement effects highlight the need for additional methodological research in the study of everyday family life.

Keywords: diary designs, parent–child conflict, Parent–child warmth, measurement reactivity, participant fatigue

Supplemental materials: <http://dx.doi.org/10.1037/dev0000081.supp>

Parent–child relationships characterized by high levels of conflict and low levels of warmth and support are associated with a host of social, psychological, and health risks for children (Repetti, Robles, & Reynolds, 2011). Although serious parent–child conflict is rare, studying experiences with more common and milder forms of tension that arise in everyday life may be important for understanding how patterns of emotion regulation develop (Sears, Repetti, Reynolds, & Sperling, 2014). These patterns increasingly solidify across the adolescent transition, setting the stage for long-term mental and physical health outcomes (Repetti, Robles, & Reynolds, 2014). Adolescence also marks a key transition in the parent–child relationship: Closeness may begin to wane and disagreements can become more salient as roles are renegotiated to

align with maturational changes (Laursen & Collins, 2009). Daily diary methods can help elucidate these relational transformations and identify the everyday processes whereby parent–child interactions both shape and accommodate adolescent well-being (e.g., Chung, Flook, & Fuligni, 2009; Lehman & Repetti, 2007; Telzer & Fuligni, 2013).

Filling a gap between single-administration surveys and laboratory observation methods, diary designs use intensive repeated sampling to assess behavior, emotions, and social interactions in families' daily lives. Compared with eliciting parent–child interactions in controlled laboratory settings, diary methods increase ecological validity. Asking family members to report on what happened *that day*, rather than retrospectively recalling weeks or months of prior experience reduces recall error compared with survey research (Hufford, 2007). Finally, diary designs allow us to examine predictors of within-family fluctuations in parent–child conflict and warmth across days, and to test whether these fluctuations and associations are moderated by more stable individual- or family-level characteristics (Repetti, Reynolds, & Sears, 2015).

At the same time, diary methods carry unique challenges. A comparatively heavy burden raises concerns about compliance and fatigue, and intensive repeated assessments risk sensitizing family members to study constructs (Barta, Tennen, & Litt, 2012). We explore these methodological issues using 56 consecutive days of diary data collected from 8- to 13-year-old children and their parents as part of a larger study of family life and health (Robles, Reynolds, Repetti, & Chung, 2013). Our design involving multiple family members and a long diary period affords a unique opportunity to evaluate burden and fatigue, as well as measurement reactivity in diary reports of parent–child conflict and warmth at

This article was published Online First December 21, 2015.

Bridget M. Reynolds, Theodore F. Robles, and Rena L. Repetti, Department of Psychology, University of California, Los Angeles.

Bridget M. Reynolds is now at Department of Pediatrics, University of Rochester Medical Center.

This research was supported by a Research Grant (9333) from the William T. Grant Foundation. We thank Richard Slatcher and Gayla Margolin for their contributions to this study, Jennifer Krull for her feedback on earlier drafts of this article, as well as the graduate students, laboratory staff, and research assistants on the project for their time and effort working with families. We also thank the parents and children who participated in this project.

Correspondence concerning this article should be addressed to Bridget M. Reynolds, Department of Pediatrics, University of Rochester Medical Center, 601 Elmwood Avenue, Box 671, Rochester, NY 14642. E-mail: Bridget_Reynolds@URMC.Rochester.edu

the beginning of the adolescent transition. Understanding how families respond to intensive repeated measurement can help improve design and maximize the benefits of diary methods in the study of child and adolescent development.

Measurement Reactivity

Measurement reactivity is broadly defined as the degree to which data are influenced by the sheer act of assessment, or as “systematically biasing effects of instrumentation and procedures on the validity of one’s data” (Barta et al., 2012, p. 108). Although a threat in most social and behavioral research, measurement reactivity is of particular concern in diary studies because of the intensity of repeated assessment. For example, asking a mother to report how much she hugs her child each day might prime her behavior and increase her likelihood of doing so the next day. In a similar way, calling attention to conflict at home through repeated tracking could enhance awareness of circumstances triggering negative interactions and thus prompt behavior change. More generally, the potential for measurement reactivity is related to the benefits of self-monitoring, in which mindful observation and tracking of one’s behavior serves as a useful therapeutic intervention (Korotitsch & Nelson-Gray, 1999). However, the benefits of self-monitoring depend upon several factors, including patient motivation and therapeutic demand for change, feedback based on previous reports, and instructions to record events and states *prior* to enacting a target behavior (Barta et al., 2012). Diary research, on the other hand, typically occurs outside of the therapeutic setting, does not include feedback or encouragement to review prior recordings, and often asks participants to report on experiences that have already occurred.

Although the conditions that promote measurement reactivity are minimized in daily diary research, testing for its effects is critical for ensuring the validity and generalizability of results. The ideal test involves experimentally manipulating sampling frequency to assess whether diary completion is causally related to behavior change. However, nonexperimental approaches can also provide insights into whether outcome variables change systematically as a function of time in the study. Because diary designs span the course of several days or weeks, normative behavior change is rarely anticipated in the absence of such measurement effects.

Current findings on measurement reactivity in diary research are mixed and come largely from work on adult health behavior. Some diary studies have found reactivity effects in which alcohol consumption decreases across days (Johnson et al., 2009; Magneberg, 1998), though this has not been universally reported (cf. Litt, Cooney, & Morse, 1998; Maas, Hietbrink, Rinck, & Keijsers, 2013). Other work shows reduced snacking and caloric intake among adults tracking eating behavior (Kirkpatrick et al., 2012; Maas et al., 2013). Although monitoring behaviors that occur habitually or outside of conscious awareness can prompt positive change, calling attention to some adverse experiences could increase the frequency or intensity of negative states.

Researchers using diaries to study negative social experiences have been particularly concerned about eliciting distress. In one of the only studies to address measurement reactivity in children, sixth grade students completed diaries about peer victimization on five random school days across a 2-week period (Nishina, 2012).

Anxiety and school aversion decreased over time, whereas physical symptoms and depressed mood showed no change. Pre- to postdiary changes in questionnaire measures of anxiety and depression did not differ between children who completed diaries and a control group of children who did not. Thus, calling children’s attention to negative peer experiences through repeated assessment did not appear to provoke distress.

In contrast, some evidence suggests negative reactivity effects with repeated assessment of romantic relationships (Merrilees, Goeke-Morey, & Cummings, 2008; Surra, Curran, & Williams, 2009). For example, husbands who completed event-contingent diaries following marital disagreements for 15 days showed declines in self-reported marital emotional quality across the study (Merrilees et al., 2008). However, behaviorally coded emotional expressions during a subsequent marital conflict-resolution task in the lab did not differ between husbands who completed the event-contingent diaries and husbands who did not. Moreover, spouses who completed once-daily diaries not contingent upon marital conflict showed no changes in ratings across days. Thus, instead of signaling true deteriorations in the marital relationship, the apparent reactive effects could be due to *response shift* (Schwartz, Sprangers, Carey, & Reed, 2004). That is, husbands whose attention was drawn to marital conflict may have recalibrated their standards for judging marital quality across study days. In such case, the observed declines could have reflected a shift toward more accurate or realistic descriptions of the relationship over time.

It remains unclear whether reactivity would emerge in daily diary research focused on conflict and warmth in parent–child interactions, spanning several weeks, and when multiple family members participate. We address this gap by testing for systematic changes in reports of parent–child conflict and warmth made by mothers, fathers, and children across 56 days of diary completion. As shown in the top panel of Table 1, the self-monitoring literature suggests that awareness gained through behavioral tracking may prompt positive change, indicated by declines in parent–child conflict and increases in parent–child warmth ratings with repeated exposure to diary measures.

Enduring parent–child relationship qualities may help predict who is most susceptible to these reactivity effects. Self-monitoring is thought to promote behavior change by highlighting discrepancies between one’s current status and a perceived ideal (Korotitsch & Nelson-Gray, 1999). Parent–child dyads with lower quality relationships may experience greater discrepancies between their actual and ideal interactions and, therefore, more self-monitoring cues to modify behavior. Thus, we might expect the sharpest declines in daily reports of conflict and the most pronounced increases in ratings of warmth among parents and children who, prior to the start of diary completion, described their relationships as more conflictual or less positively involved.

However, the studies previously reviewed suggest that, if evident, reactive effects will likely be small and perhaps even in directions counter to self-monitoring predictions. As observed among husbands tracking episodes of marital disagreement (Merrilees et al., 2008), parent and child ratings of positive relationship qualities may decline as diaries call attention to more conflictual aspects of daily interactions. Participation of multiple family members affords a unique opportunity to test whether any changes in diary reports over time reflect increasingly accurate or realistic

Table 1
Conceptual Definitions and Hypotheses Associated With Measurement Reactivity and Fatigue

Construct	Conceptual definition	Hypotheses
Measurement reactivity	Exposure to repeated measurement produces systematic change in the constructs under investigation. Enhanced awareness gained through self-monitoring may prompt behavioral improvements over time.	<p>Descriptions of more positive interactions over time:</p> <ul style="list-style-type: none"> ■ Decreases in daily reports of parent-child conflict ■ Increases in daily ratings of parent-child warmth ■ More pronounced changes among those reporting lower-quality relationships at baseline <p>Improvements in response accuracy over time:</p> <ul style="list-style-type: none"> ■ Strengthening agreement between parent and child ratings of daily interactions ■ Less mood-congruent retrieval over time: negative mood less strongly predicts same-diary ratings of parent-child interactions
Fatigue	Participants become tired, bored, or overburdened by measurement procedures, resulting in deteriorations in the amount or quality of data provided over time.	<p>Reductions in amount or timeliness of data provided:</p> <ul style="list-style-type: none"> ■ Declines in daily diary completion and on-time compliance over time <p>Deteriorations in response accuracy over time:</p> <ul style="list-style-type: none"> ■ Weakening agreement between parent and child ratings of daily interactions ■ More mood-congruent retrieval over time: negative mood more strongly predicts same-diary ratings of parent-child interactions

descriptions of the parent-child relationship. If so, we might expect strengthening agreement or convergence between same-day parent and child ratings of their interactions with repeated exposure to the diary measures.

Calling attention to parent-child interactions might also improve accuracy by facilitating recall. With repeated assessments, responses might become increasingly independent of extraneous factors such as participant mood at the time of diary completion. Thus, another way to capture improvements in accuracy is to test for a declining influence of mood-congruent retrieval (Kihlstrom, Eich, Sandbrand, & Tobias, 1999), inferred through weakening within-person associations between negative mood and same-diary ratings of parent-child conflict or warmth across time in the study. Although previous research has linked daily reports of negative mood with descriptions of fewer positive and more negative parent-child interactions (Flook, 2011; Kiang & Buchanan, 2014), no systematic change in the strength of these associations would be anticipated absent such measurement-related influences. In this context, however, we must also consider the potential for participant fatigue, another consequence of the burden of repeated measurement that has implications for response accuracy.

Participant Burden and Fatigue

Burden and fatigue are major concerns in diary research, especially in studies that involve children. Fatigue occurs when participants become bored or overburdened by repeated assessments and skip diaries, complete entries late, or put less attention into responding over time (Hufford, 2007). Fatigue can introduce noise, reduce accuracy, and lead to bias if data are systematically missing or noncompliant. For example, an increased likelihood of skipping diaries on weekends than on weekdays is problematic for statistical models that assume data are missing at random unless weekday is modeled as a covariate (Schafer & Graham, 2002).

Traditionally, researchers have addressed burden and fatigue by reporting how many diaries participants completed out of those expected. Beyond the number of diaries submitted, a more stringent definition of compliance includes assessing diary completion dates and times. Although creative experimental work has questioned the validity of self-reports of this information (Broderick, Schwartz, Shiffman, Hufford, & Stone, 2003), compliance in studies using electronic date-time stamping is generally good, with on-time completion rates around 90% among adults (Hufford, 2007; Shiffman, Stone, & Hufford, 2008) and between 80 and 85% in several 14-day studies of high school students (Chung et al., 2009; Kiang & Buchanan, 2014). Little is known about compliance among younger children and across longer diary periods. Moreover, although several daily diary studies with families have included both parent and child reports (Lehman & Repetti, 2007; Timmons & Margolin, 2015), more work is needed to clarify how participation of multiple family members might impact compliance rates.

As shown in the bottom panel of Table 1, this study addresses burden and fatigue by testing for within-person declines in diary completion and on-time compliance rates among mothers, fathers, and children across 8 weeks of study participation. Our design involving multiple family members also allows for a more nuanced assessment of fatigue effects. Although difficult to test directly, fatigue-related declines in accuracy can be inferred through weakening agreement between parent and child reports about their daily interactions across the study. Of course, such a finding would not pinpoint whether fatigue biased the descriptions provided by parents, children, or both, and would thus need to be considered in light of additional evidence.

As previously noted, another way to capture accuracy is to examine how strongly participants' emotional states predict their diary ratings. If fatigue leads to less careful or attentive reporting, then mood-congruent retrieval may increasingly influence participants' responses to diary items over time (Kihlstrom et al., 1999).

Specifically, within-person links between negative mood and same-diary ratings of parent–child conflict or warmth might strengthen across time if overburdened participants put less effort into recalling details of the day’s interactions. If this is true among children but not parents, and if agreement between parent and child reports weakens over time, then we could infer child fatigue with greater confidence.

The Current Study

We address measurement reactivity and fatigue using data collected from 8- to 13-year-old children and their parents as part of a larger 56-day diary study of family stress and susceptibility to upper respiratory infection. Our lengthy diary period—required to capture the onset of cold symptoms and to reliably sample a range of family experiences—provided a unique opportunity to assess participant burden and fatigue. Moreover, the simultaneous involvement of parents and children from two-parent households allowed us to separately test for measurement-related change in descriptions of conflict and warmth in the mother–child and father–child relationship from the perspective of multiple reporters. Although we did not expect these constructs to be particularly reactive in comparison with other aspects of the parent–child relationship, we chose to focus on conflict and warmth, given their developmental importance, in general (Laursen & Collins, 2009; Repetti et al., 2011), and interest in their assessment via diary methods, in particular (Chung et al., 2009; Lehman & Repetti, 2007; Repetti, 1996; Telzer & Fuligni, 2013; Timmons & Margolin, 2015). Our specific hypotheses are outlined in Table 1. Although we anticipate some evidence consistent with both reactivity and fatigue, ideally these effects will be small and addressable through statistical controls and design improvements.

Method

Participants

Two-parent families were recruited in the Los Angeles area through flyers distributed in schools, community centers, and medical clinics; newspaper advertisements; and direct mailings based on a youth marketing address list. At least one parent and a target child between the ages of 8 and 13¹ were screened for a variety of mental and physical health problems to yield a generally healthy sample free of major chronic illness. Data were collected in three yearly cohorts between the months of September and May. Although both parents were encouraged to take part in the study, we only required the participation of one. Our sample of 47 families includes 47 mothers (M age = 43.29, SD = 6.31), 39 fathers (M age = 43.67, SD = 8.10), and 47 target children (19 boys, 28 girls; M age = 11.28, SD = 1.50). No differences emerged between families that included a participating father and families that did not on any of the child- or mother-report variables used in this study. Parents reported a median personal income within a \$31,850 to \$82,400 bracket, with 57% of mothers (59% of fathers) having attained a 4-year college degree or higher. Roughly 45% of mothers worked full time (80% of fathers), 21% worked part time (13% of fathers), and the remainder reported being homemakers, unemployed, disabled, or “other.” According to self-reports, parents were 45% non-Hispanic White, 22% Latino/His-

panic, 17.5% African American, 12.5% Asian, 1.5% Native American, and 1.5% “Other.”

Procedure

During an initial visit that typically occurred in the family’s home, researchers discussed study procedures with the family, answered questions, and obtained informed consent from parents and assent from children (approved by the UCLA Institutional Review Board). Participants were then interviewed separately. Parents completed a series of online baseline questionnaires at their convenience prior to the start of the diary phase of the study. Parents and children were trained on diary procedures during a second home visit, typically within a week of the initial visit. Children completed online questionnaires at the end of the second visit, which enabled a researcher to answer questions and ensure that children were responding independently of their parents. The 8-week daily diary began the first Saturday following the second visit.

Family members were asked to complete their diaries online before bed each night by logging into our study web portal. Personalized “home” pages allowed communication with lab staff through private messages, displayed study reminders, and provided a link to the current-day online diary (blocks of items were randomly ordered across days of the week). Compliance was monitored via automated date–time stamping on SurveyMonkey.com. If a participant failed to complete the online diary on three or more consecutive days, research staff called the family to troubleshoot compliance issues. Although not a condition for eligibility, all families had home Internet access. In case of technical difficulties, each participant was given 2 weeks’ worth of paper diaries and a tamper-resistant electronic date–time stamp device, with instructions to stamp the diary immediately upon completion, place it in a postage-paid envelope and stamp across the seal, and then mail it the following morning. Each parent and child earned \$20 and \$10, respectively, per week of participation in the diary phase, in addition to a \$5 bonus gift card per week of 100% compliance (i.e., all diaries completed on the evening due or before 9:00 a.m. the next morning).

Although not a focus of the current study, other data collection procedures may have acted as sources of burden. In addition to evening surveys, the diary phase involved completion of sleep diaries each morning and, during Weeks 3 and 6, saliva samples were collected on 4 consecutive days to assess diurnal cortisol (5 per day \times 8 days total). Additionally, cold symptoms were monitored on a daily basis, and suspected upper respiratory infections were verified within 48 hr via an in-home visit by a pair of trained clinicians (see Robles et al., 2013). Following the daily diary phase, families attended a final laboratory visit that involved the collection of blood samples, the child’s participation in a stressor task, and a feedback interview.

¹ We recruited families of children aged 8 to 12 years for our first cohort, and 9 to 13 years for the remaining two cohorts. Forty of the 47 target children were between 10 to 13 years of age.

Table 2

Descriptive Statistics for Child, Mother, and Father Reports on Daily Diary Scales and Baseline Questionnaire Measures

Construct and reporter	Daily diary							Baseline questionnaire (PEQ)				
	N_t	R_{KF}	R_C	ICC	M	SD	Range	N	α	M	SD	Range
Mother-child conflict												
Child report	2392	.99	.74	.31	1.16	.34	1.00–3.00	46	.91	1.57	.54	1.00–3.08
Mother report	2418	.99	.83	.30	1.17	.27	1.00–3.00	47	.92	1.63	.56	1.00–3.17
Father-child conflict												
Child report	2209	.99	.74	.22	1.13	.30	1.00–3.00	46	.81	1.60	.42	1.00–2.92
Father report	1830	.99	.80	.20	1.11	.21	1.00–2.67	39	.84	1.45	.35	1.00–2.08
Mother-child warmth												
Child report	2391	1.00	.72	.59	2.51	.51	1.00–3.00	46	.79	3.62	.41	2.58–4.00
Mother report	2418	.99	.68	.50	2.18	.47	1.00–3.00	47	.79	3.71	.33	2.50–4.00
Father-child warmth												
Child report	2211	1.00	.71	.57	2.44	.52	1.00–3.00	46	.79	3.40	.46	2.42–4.00
Father report	1827	1.00	.69	.53	1.94	.45	1.00–3.00	39	.89	3.65	.40	2.58–4.00
Negative mood												
Child report	2470	1.00	.72	.54	1.26	.38	1.00–3.33					
Mother report	2485	1.00	.83	.40	1.45	.51	1.00–4.00					
Father report	1935	1.00	.85	.51	1.35	.46	1.00–3.88					

Note. Daily diary descriptive statistics were calculated across time and participants (N_t). R_{KF} is an estimate of between-person reliability and is calculated across participants and days, whereas R_C is an estimate of within-person reliability (Cranford et al., 2006). Parallel parent and child versions of the PEQ (Elkins, McGue, & Iacono, 1997) measured global perceptions of conflict and warmth/positive involvement in the parent-child relationship at baseline. ICC = intraclass correlation coefficient; PEQ = Parental Environment Questionnaire.

Measures

Items composing the child and parent daily diary scales can be found in the supplemental online materials (SOM, Table A). Summary statistics for all measures are presented in Table 2.

Child daily diary. Child diaries included items based on the Youth Everyday Social Interaction and Mood measure (YES-I-a.m.; Repetti, 1996) to assess the parent-child relationship and mood. Events at school, perceptions of the marital relationship, cold symptoms, and health behaviors were also assessed. The present study focuses on perceived parent-child warmth (three items; e.g., “My mom gave me love and attention today”) and parent-child conflict (three items; e.g., “My dad got mad at me today”). Each item was rated separately for mothers and fathers on a 3-point scale: 1 = *not at all*, 2 = *some*, 3 = *a lot*. Negative mood was assessed with six items (e.g., “I was sad,” “I felt worried,” “I was angry”) on a 4-point response scale: 1 = *not at all*, 2 = *some of the day*, 3 = *most of the day*, 4 = *all day*. Responses were averaged to create scale scores each day. On average, children completed the online diary in 4.41 min ($SD = 3.14$).

Parent daily diary. Items drawn from the Parent Home Data Questionnaire (Dumas, Margolin, & John, 2003) assessed the parent-child relationship and mood. Parents were also asked to report on marital interactions, child disposition, cold symptoms, and health behaviors. For the present study, nine items assessed conflict (e.g., “Today I nagged my child”) and four assessed warmth (e.g., “Today I praised my child”) between the parent and the target child. Response scales were identical to those used for the child diary. Parent negative mood was assessed with eight items (e.g., “sad,” “on edge,” “hostile”), rated on a 4-point scale indicating how accurately each described the parent’s feelings that day (1 = *completely inaccurate*, 2 = *mostly inaccurate*, 3 = *mostly accurate*, 4 = *completely accurate*). Responses were averaged to create scale scores each day. On average, parents completed the online diary in 7.60 min ($SD = 4.24$).

Between- and within-person reliabilities for daily diary scales, shown in Table 2, were estimated using a generalizability theory framework (Cranford et al., 2006). The between-person estimates (R_{KF}), consistently .99 or above, indicate that the diary scales measured relatively stable individual differences with excellent reliability, given the diary period (R_{KF} subscripts refer to averaging across $K = 56$ fixed [F] time points). The within-person estimates (R_C), ranging from .68 to .85, indicate that diary scales were able to reliably detect meaningful day-to-day change (C) within participants across all 56 days. R_C values were generally higher for parent diary scales, which included more items compared with child measures.

Global parent-child relationship qualities. To assess global perceptions of the parent-child relationship at baseline, children and parents completed parallel versions of the Parental Environment Questionnaire (PEQ; Elkins, McGue, & Iacono, 1997), including two 12-item scales: Conflict (e.g., parent version: “My child and I often argue”; child version: “My mother and I often get into arguments”) and Positive Involvement (e.g., “My child shares concerns with me”; “I share concerns with my mother”). Each parent completed the measure in reference to the target child. Children completed two versions to assess conflict and Involvement with mothers and with fathers. Response options ranged from 1 (*definitely false*) to 4 (*definitely true*), and ratings were averaged; higher Conflict scores indicate more conflict and higher Involvement scores reflect greater warmth and positive Involvement. The PEQ has adequate validity (Elkins et al., 1997), and internal consistency was acceptable in the current study (see Table 2).

Results

We present results in three sections. The section on *participant burden and fatigue* describes diary completion and on-time compliance rates across the study. *Measurement reactivity* is then addressed by testing for systematic changes in daily reports of

conflict and warmth over time and by investigating whether trajectories of change are moderated by baseline questionnaire ratings of parent–child relationship qualities. We then examine implications of measurement reactivity and fatigue for diary *response accuracy* by testing for temporal changes in (a) agreement between parent and child reports of their daily interactions, and (b) the strength of within-person, bivariate associations between negative mood and same-day ratings of parent–child conflict and warmth. Examples of the multilevel equations tested in each section can be found in SOM Table B.

Participant Burden and Fatigue

A total of 7,029 daily diaries were completed across all participants (4,549 by parents, 2,480 by children), of which 6,897 (98%) were compliant, meaning that they were completed on the evening due or before 9:00 a.m. the next morning. On average, mothers completed 95% of the 56 expected diaries, whereas fathers and target children completed 94% of the diaries (see top panel of Table 3). The vast majority of diaries were completed online (95%).

To test the hypotheses that diary completion and on-time compliance will decline across study weeks, we estimated logistic multilevel models using SAS PROC GLIMMIX. Three-level random intercept models assumed diaries (Level 1) were nested within individuals (Level 2), who were in turn nested within families (Level 3).² In addition to testing for linear trends in rates of diary completion across study weeks (coded 0 to 7, with Week 1 = 0), we also included a dummy variable modeling differences in the probability of diary completion on weekdays (coded 1) versus weekend days (coded 0) and an interaction term (Week × Weekday) testing whether this effect was stable across weeks. Results are shown at the bottom of Table 3. Although the probability of diary completion on weekend days remained stable over time, a significant Week × Weekday interaction indicated that the probability of diary completion on weekdays dropped slightly, from a predicted probability of .99 at the start to .95 at the end of the 8 weeks. No significant time trends emerged in models of on-time diary compliance. Results did not change when compliance was coded more strictly, with a cutoff of 4:00 a.m. rather than 9:00 am the next day (SOM, Table C).

Measurement Reactivity

We next addressed measurement reactivity by estimating trajectories of change in diary ratings of parent–child conflict and warmth separately for mothers, fathers, and target children using SAS PROC MIXED. Two-level models assumed diary days (Level 1) were nested within individuals (Level 2), and a first-order autoregressive within-person error structure (AR[1]) accounted for time dependencies in the daily reports. For each dependent variable, we stepped through a taxonomy of models (see SOM, Table D), starting with an empty model to calculate the intraclass correlation coefficients in Table 2 (ICC; the proportion of total outcome variation attributable to differences between [vs. within] individuals). We then estimated unconditional growth models that included time in study as the sole predictor (*Day* coded 0 to 55, first day = 0).³ Models with curvilinear effects were also examined, but quadratic terms did not significantly predict any of the

outcomes. Because prior research has shown that supportive and conflictual parent–child interactions are more likely to occur on days parents spend more time with their children (Almeida, Wethington, & McDonald, 2001), we next added a dummy variable to test for differences in reports made on weekdays (coded 1) versus weekend days (coded 0), as well as an interaction term testing whether trajectories of change differed across weekdays versus weekend days of the study.⁴ At this step, the within-person (Level 1) equation was as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Day}_{ij}) + \beta_{2j}(\text{Weekday}_{ij}) + \beta_{3j}(\text{Day} * \text{Weekday}_{ij}) + e_{ij} \quad (1)$$

where Y_{ij} , for example, is the level of mother–child warmth reported on day i by child j . To examine whether change trajectories vary as a function of global parent–child relationship qualities, same-reporter PEQ Conflict or Involvement ratings were then separately modeled as between-person, Level 2 predictors of the intercept (β_{0j}) and all slope coefficients shown above:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{PEQ}_j) + u_{0j} & \beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{PEQ}_j) \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{PEQ}_j) + u_{1j} & \beta_{3j} &= \gamma_{30} + \gamma_{31}(\text{PEQ}_j) \end{aligned} \quad (2)$$

where PEQ_j , for example, is child j 's PEQ Involvement score, centered about the grand mean for all children in the sample. Below, we present results for unconditional growth models and for models incorporating weekend-weekday differences. We then turn our attention to the moderated models testing for differences in temporal trends as a function of global relationship qualities.

Parent-child conflict change trajectories. In unconditional growth models, child reports of conflict with parents showed no linear change over time, whereas both mother ($\gamma = -.002$, $SE = .001$, $p < .01$) and father ($\gamma = -.001$, $SE = .0003$, $p < .001$) reports of conflict with the target child showed a linear decrease across days. To characterize the magnitude of the linear effect of time on mother and father reports of parent–child conflict, pseudo- R^2 statistics were calculated as the percent reduction in explainable within-person residual variance from the empty model to the unconditional growth model (Singer & Willett, 2003). The residual variance for mother-reported conflict decreased from .0540 to .0516, representing a 4.4% reduction accounted for by study day. For fathers, residual variance showed a 1.5% reduction (.0341 to .0336). Using Cohen's (1988) guidelines, the linear effect of study day on parent reports of conflict is small.

We next tested whether conflict ratings and trajectories of change differed on weekdays compared with weekend days across the study. As shown in Table 4, Model A, family members tended to report less parent–child conflict on weekdays than on weekend days (γ_{20}), though this difference did not reach statistical significance for child-reported conflict with mothers. A significant

² Diaries are technically crossed, rather than nested, with the individual and family levels. The correlation between family members' intercepts is estimated as a variance in the three-level model and, therefore, constrained to be positive.

³ Trajectories were allowed to vary randomly across individuals except in models of father-reported conflict, for which this specification led to convergence problems.

⁴ Random effects associated with the weekday and interaction terms were not significant in preliminary analyses and were dropped in accordance with recommendations (Nezlek, 2003).

Table 3

Descriptive Statistics for Overall Daily Diary Completion and Compliance by Type of Family Member (Top) and Results of Multilevel Logistic Models Predicting Probability of Diary Completion and Compliance Over Time (Bottom)

Descriptive statistics	Completed diaries				Compliant diaries			
	<i>N</i>	Mean (<i>SD</i>)	Range	% of expected	Mean (<i>SD</i>)	Range	% of expected	% of completed
Mother	47	53.17 (6.43)	28–61	95	52.26 (7.26)	28–61	93	98
Father	39	52.56 (7.87)	21–60	94	51.46 (8.74)	21–60	92	98
Target child	47	52.77 (6.64)	28–58	94	51.79 (7.53)	24–58	92	98

Logistic MLM results	Completed diaries			Compliant diaries		
	Estimate	(<i>SE</i>)	<i>p</i> value	Estimate	(<i>SE</i>)	<i>p</i> value
Intercept	3.906	(.32)	<.001	4.840	(.40)	<.001
Week	-.035	(.04)	.403	-.001	(.07)	.995
Weekday	1.000	(.24)	<.001	.298	(.37)	.418
Week × Weekday	-.235	(.05)	<.001	-.002	(.09)	.984

Note. Diaries were compliant if done the evening due or before 9:00 a.m. the next day. The ranges in the top panel show that a number of participants (7 mothers, 3 fathers, 6 children) completed additional diaries beyond the 56 required days. % of expected indicates the proportion of the 56 expected diaries represented by the means, while % of completed indicates the proportion of the mean number of completed diaries that were also compliant. The bottom panel presents parameters and SEs from two separate logistic multilevel models predicting log odds of diary completion (left) and on-time compliance (right) across study weeks, with random intercepts and fixed slopes. Diaries (Level 1) are nested within individuals (Level 2), who are in turn nested within families (Level 3). *Week* represents week of study (coded 0–7; Week 1 = 0) and *weekday* is a dummy code (weekend = 0, weekday = 1). MLM = multilevel model; *SE* = standard error.

Day × Weekday interaction (γ_{30}) also emerged for parent descriptions of conflict. Examination of simple slopes revealed that weekend reports showed a greater decline across study days (mother: $\gamma = -.003$, $SE = .001$, $p < .001$; father: $\gamma = -.003$, $SE = .001$, $p < .001$) than did weekday reports (mother: $\gamma = -.001$, $SE = .0005$, $p < .05$; father: $\gamma = -.001$, $SE = .0003$, $p < .05$). As shown in Figures 1A and 1B, this resulted in a narrowing of the initial difference in ratings made on weekdays versus weekend days (i.e., the weekend “bumps” in parent reports of conflict flatten out over time). According to a calculator that uses the Johnson-Neyman technique to obtain regions of significance for multilevel interactions (Preacher, Curran, & Bauer, 2006), the difference between weekday and weekend reports of parent-child conflict was only significant during the first 3 weeks of the study for mother reports and the first 4 weeks for father reports. Pseudo- R^2 statistics (.001 to .02) indicate that the addition of weekday and its interaction with study day resulted in negligible to small reductions in within-person variance compared with unconditional growth models for all family members. In sum, declines in parent-reported conflict were in line with predictions based on the self-monitoring literature, though effects were small.

Parent-child warmth change trajectories. Unconditional growth models revealed that child reports of warmth did not change over time, whereas both mother ($\gamma = -.003$, $SE = .001$, $p < .01$) and father ($\gamma = -.002$, $SE = .001$, $p < .01$) reports showed a linear decline across days. Pseudo- R^2 statistics indicated that the linear effect of time in study on parent reports of warmth was between small and medium (.04 to .07). Table 5, Model A, shows that all family members initially reported less warmth on weekdays compared with weekend days; however, this effect was qualified by a Day × Weekday interaction for child and father reports (see Figures 1C and 1D). An examination of simple slopes revealed that fathers’ weekend reports of parent-child warmth decreased over time ($\gamma = -.004$, $SE = .001$, $p < .001$), whereas their weekday reports showed no linear change across the study ($\gamma = -.001$, $SE = .001$, *ns*). Children’s

weekend reports of father-child warmth decreased ($\gamma = -.001$, $SE = .001$, *ns*) and their weekday reports increased across the study ($\gamma = .001$, $SE = .001$, *ns*); however, neither simple slope was statistically significant. Regions of significance indicated that the difference between weekday and weekend reports of warmth remained significant until the last week of the study for both children and fathers. Pseudo- R^2 statistics indicate that the combined effect of weekday and its interaction with study day was small for child and mother reports (.01 to .03), and between small and medium for father ratings of parent-child warmth (.08).

Taken together, the results suggest possible measurement reactivity in parent reports of conflict and warmth, although the observed declines in warmth were not in line with hypotheses. Both parents’ weekend ratings of conflict, and fathers’ weekend ratings of warmth, showed greater declines across the study compared with reports made on weekdays. There was no evidence of reactivity in children’s reports of conflict, but the weekend-to-weekday difference in their reports of father-child warmth narrowed across the study. Findings remained the same when on-time diary compliance was controlled as a time-varying covariate in all models (SOM, Table E).

Moderated change trajectories. Next, we examined whether change trajectories might vary across families as a function of global parent-child relationship qualities measured via one-time survey at baseline. We expected sharper improvements in ratings of daily interactions over time among parents and children who perceived lower overall levels of Involvement or higher overall levels of Conflict in their relationship prior to the start of diary completion.

Results of models testing parent-child Involvement as a moderator of change are shown in SOM Table F. Baseline PEQ Involvement ratings moderated trajectories of change in mothers’ daily reports of both parent-child conflict and warmth (see Figures 2A and 2C). An examination of simple slopes revealed no linear change over time in daily reports of conflict among mothers higher in Involvement (+1SD; $\gamma_{10} = -.001$, $SE = .001$, *ns*), but a

Table 4

Multilevel Models of Daily Parent–Child Conflict as a Function of Study Day, Weekday Versus Weekend, and Same-Day, Same-Reporter Negative Mood

Parameter		Mother-Child Conflict				Father-Child Conflict			
		Child-Report		Mother-Report		Child-Report		Father-Report	
		Model A	Model B	Model A	Model B	Model A	Model B	Model A	Model B
Fixed									
Intercept	γ_{00} (SE)	1.21 (.03)***	1.21 (.03)***	1.27 (.03)***	1.28 (.03)***	1.17 (.02)***	1.17 (.02)***	1.20 (.02)***	1.20 (.02)***
Day	γ_{10} (SE)	-.001 (.001)	-.001 (.001)	-.003 (.001)***	-.003 (.001)***	-.001 (.001)	-.001 (.001)	-.003 (.001)***	-.002 (.001)
Weekday	γ_{20} (SE)	-.03 (.02)	-.02 (.02)	-.07 (.02)***	-.08 (.02)***	-.05 (.02)*	-.04 (.02)*	-.08 (.02)***	-.08 (.02)***
Day × Weekday	γ_{30} (SE)	.0003 (.001)	-.0001 (.001)	.002 (.001)**	.002 (.001)**	.001 (.001)	.001 (.001)	.002 (.001)**	.002 (.001)**
Neg mood	γ_{40} (SE)		.24 (.08)**		.19 (.05)***		.21 (.07)**		.03 (.05)
Neg Mood × Day	γ_{50} (SE)		.006 (.002)*		-.002 (.001)		.005 (.002)*		.002 (.002)
Neg Mood × Weekday	γ_{60} (SE)		.17 (.08)*		-.05 (.05)		.04 (.08)		.04 (.06)
Neg Mood × Day × Weekday	γ_{70} (SE)		-.007 (.003)**		.002 (.002)		-.006 (.003)*		-.002 (.002)
Random									
Intercept	σ_0^2	.030***	.035***	.032***	.036***	.010**	.012***	.008***	.008***
Day slope	σ_1^2	.00001	.000003	.00001**	.00001**	.00001*	.000003		
Mood slope	σ_3^2		.062**		.029***		.039**		
Covariance									
Intercept-Day	σ_{01}	.0001	-.00001	-.0003*	-.0003**	.0001	.0001		
Intercept-Mood	σ_{04}		.009		-.001		.010		
Day-Mood	σ_{14}		-.0001		.0002*		.00001		
Residual	σ_2^2	.079***	.065***	.051***	.047***	.068***	.059***	.033***	.032***
AR(1)	ρ	.10***	.10***	.09***	.08***	.12***	.12***	.15***	.14***

Note. The outcome and reporter for each model is listed in the column headings. *Day* is a linear growth term representing day in the study (coded 0–55, first day = 0) and *Weekday* is a dummy code (weekday = 1, weekend day = 0). Daily negative mood (*Neg mood*) is reported by the same person making the outcome rating, on the same day, and is centered about each individual's mean score. *AR(1)* is the autocorrelation.

* $p < .05$. ** $p < .01$. *** $p < .001$.

significant decline across days for mothers lower in Involvement ($-1 SD$; $\gamma_{10} = -.005$, $SE = .001$, $p < .001$). The opposite pattern emerged in daily reports of warmth: mothers higher in Involvement ($+1 SD$) showed a significant linear decline across days ($\gamma_{10} = -.006$, $SE = .002$, $p < .001$), whereas mothers lower in Involvement ($-1 SD$) showed no linear change over time ($\gamma_{10} = -.0004$, $SE = .002$, *ns*).

Among fathers, higher Involvement ratings were associated with a larger weekend-to-weekday difference in daily reports of both conflict and warmth (see Figures 2B and 2D; SOM Table F). However, this effect was qualified by a significant Day × Weekday × Involvement interaction in predicting conflict. Involvement did not moderate linear change in fathers' weekday ratings of conflict, but did predict change in fathers' weekend reports ($\gamma_{11} = -.003$, $SE = .001$, $p < .05$). An examination of simple trajectories of change for weekend diaries revealed a steeper decline in conflict ratings among fathers higher in Involvement ($+1 SD$; $\gamma_{10} = -.004$, $SE = .001$, $p < .001$) compared with fathers lower in Involvement ($-1 SD$; $\gamma_{10} = -.002$, $SE = .001$, $p < .05$).

Children's baseline perceptions of parent–child Involvement did not moderate any temporal trends in their diary reports of conflict or warmth with either parent. Furthermore, baseline perceptions of parent–child Conflict did not moderate time-related changes in any family member's daily reports of conflict or warmth (SOM Table G).

Overall, our results suggest that parents' baseline perceptions of positive Involvement predict temporal change in daily reports of warmth and conflict, though patterns of moderation differed between mothers and fathers. We expected greater improvements in ratings of daily interactions among those who began the study

reporting overall lower quality parent–child relationships. Sharper declines in daily conflict ratings among less involved mothers were in line with this hypothesis; however, greater decreases in daily reports of warmth and conflict among *more* involved mothers and fathers, respectively, were contrary to predictions.

Response Accuracy

We next addressed the implications of reactivity and fatigue for the accuracy of diary responses over time. We operationalized improvements in accuracy two ways: (a) strengthening agreement between same-day parent and child reports, and (b) weakening associations between negative mood and same-diary ratings of parent–child interactions across days. We predicted improved accuracy in the context of reactivity, but reduced accuracy in the context of fatigue.

Agreement between parent and child daily reports. We first tested whether agreement between parent and child reports of daily interactions changed as a function of time in the study. We examined models in which parent reports predicted same-day child report outcomes and vice versa. For example, child-reported conflict with mother was modeled with the following Level 1 equation, where X_{ij} is the same-day mother rating of conflict, centered about the mother's mean:

$$\begin{aligned}
 Y_{ij} = & \beta_{0j} + \beta_{1j}(\text{Day}_{ij}) + \beta_{2j}(\text{Weekday}_{ij}) + \beta_{3j}(\text{Day} * \text{Weekday}_{ij}) \\
 & + \beta_{4j}(X_{ij}) + \beta_{5j}(\text{Day} * X_{ij}) + \beta_{6j}(\text{Weekday} * X_{ij}) \\
 & + \beta_{7j}(\text{Day} * \text{Weekday} * X_{ij}) + e_{ij}
 \end{aligned}
 \quad (3)$$

Table 5

Multilevel Models of Daily Parent–Child Warmth as a Function of Study Day, Weekday Versus Weekend, and Same-Day, Same-Reporter Negative Mood

Parameter	Mother-Child Warmth				Father-Child Warmth				
	Child-Report		Mother-Report		Child-Report		Father-Report		
	Model A	Model B	Model A	Model B	Model A	Model B	Model A	Model B	
Fixed									
Intercept	γ_{00} (SE)	2.54 (.06)***	2.53 (.06)***	2.33 (.06)***	2.32 (.06)***	2.51 (.06)***	2.51 (.07)***	2.13 (.06)***	2.12 (.06)***
Day	γ_{10} (SE)	-.00003 (.001)	.0003 (.001)	-.003 (.001)**	-.003 (.001)***	-.001 (.001)	-.001 (.001)	-.004 (.001)***	-.004 (.001)***
Weekday	γ_{20} (SE)	-.10 (.03)***	-.10 (.03)***	-.12 (.03)***	-.11 (.03)***	-.14 (.03)***	-.14 (.03)***	-.25 (.03)***	-.23 (.03)***
Day × Weekday	γ_{30} (SE)	.002 (.001)	.002 (.001)	.001 (.001)	.001 (.001)	.002 (.001)*	.002 (.001)	.003 (.001)**	.003 (.001)**
Neg mood	γ_{40} (SE)		-.34 (.08)***		-.16 (.06)**		-.20 (.09)*		-.20 (.08)*
Neg Mood × Day	γ_{50} (SE)		.003 (.003)		.001 (.002)		-.003 (.003)		.002 (.003)
Neg Mood × Weekday	γ_{60} (SE)		.06 (.10)		.03 (.07)		-.05 (.11)		.13 (.10)
Neg Mood × Day × Weekday	γ_{70} (SE)		.001 (.003)		.0003 (.002)		.007 (.003)		-.002 (.003)
Random									
Intercept	σ_0^2	.162***	.169***	.131***	.129***	.162***	.169***	.099***	.106***
Day Slope	σ_1^2	.00002**	.00002**	.00002**	.00002**	.00002**	.00002**	.00001*	.00001*
Mood Slope	σ_3^2		.038*		.005		.022		
Covariance:									
Intercept-Day	σ_{01}	-.0004	-.0005	-.0006	-.0006	-.0004	-.0005	.0002	.0001
Intercept-Mood	σ_{04}		.008		.001		-.007		
Day-Mood	σ_{14}		-.0005*		-.0002		-.0001		
Residual	σ_2^2	.101***	.095***	.102***	.099***	.112***	.107***	.089***	.090***
AR(1)	ρ	.20***	.22***	.23***	.23***	.16***	.17***	.18***	.16***

Note. The outcome and reporter for each model is listed in the column headings. *Day* is a linear growth term representing day in the study (coded 0–55, first day = 0) and *Weekday* is a dummy code (weekday = 1, weekend day = 0). Daily negative mood (*Neg mood*) is reported by the same person making the outcome rating, on the same day, and is centered about each individual's mean score *AR(1)* is the autocorrelation.

* $p < .05$. ** $p < .01$. *** $p < .001$.

The interaction between mother-reported conflict and study day (β_{5j}) tested for systematic changes in agreement as a function of time in the study, whereas the three-way interaction (β_{7j}) allowed rates of change in agreement to differ across weekday versus weekend diary reports. We then tested a similar model with mother conflict ratings serving as the outcome instead of the predictor. Slopes associated with parent–child interaction ratings were allowed to vary randomly across individuals in all models. Results can be found in SOM Tables H and I.

Agreement between mother and child reports of conflict declined across the study. Child conflict ratings predicted same-day mother-reported conflict on the first diary day ($\gamma_{40} = .44$, $SE = .06$, $p < .001$), but an interaction with study day ($\gamma_{50} = -.004$, $SE = .002$, $p < .05$) indicated that the strength of this association declined, on average, by .004 with each passing day of participation. Similar results were obtained when mother reports of conflict were modeled as a predictor rather than the outcome: Mothers' initial ratings predicted same-day child reports of conflict ($\gamma_{40} = .56$, $SE = .09$, $p < .001$), and the strength of this association declined by a similar amount over time ($\gamma = -.005$, $SE = .002$, $p < .01$), but only across weekday diaries. Mother weekend ratings continued to predict child-reported conflict at the same strength across the entire study (i.e., a significant Conflict × Day × Weekday interaction). Agreement between child and father reports of conflict remained stable over time, as did agreement between parent and child daily reports of warmth.

Bivariate associations with negative mood. Prior to testing bivariate models, we first examined temporal trends in daily reports of negative mood and found no significant linear change as a function of study day. Parent reports of negative mood were

significantly higher on weekdays compared with weekend days, but this effect was stable over time (SOM, Table J). Next, we tested whether within-person associations between same-day reports of negative mood and parent–child interactions changed as a function of time in the study. For example, child-reported conflict with mother was modeled according to Equation 3, but with X_{ij} now representing the child's same-day negative mood rating (centered about his or her mean negative mood score).⁵

As shown in Table 4, Model B, on days when children and mothers reported more negative mood, they also reported greater parent–child conflict (γ_{40}). Among children, however, these associations were qualified by significant two- (γ_{50}) and three-way (γ_{70}) interactions involving time in study. For reports children made on weekend days, the association between negative mood and conflict with both mothers and fathers strengthened over time (γ_{50}). To probe the significant three-way interactions (γ_{70}), we examined the same models presented in Table 4, but with weekday diaries coded as the reference category instead of weekend diaries. Whereas bivariate associations became stronger over time for children's weekend reports, no linear change was observed in associations between negative mood and child weekday reports of conflict with mothers ($\gamma_{50} = -.0017$, $SE = .002$, ns) or with fathers ($\gamma_{50} = -.0007$, $SE = .002$, ns). According to results presented in Table 5, Model B, all family members reported less

⁵ Mood slopes were allowed to vary randomly across individuals, except in father-report models, for which this specification led to convergence problems.

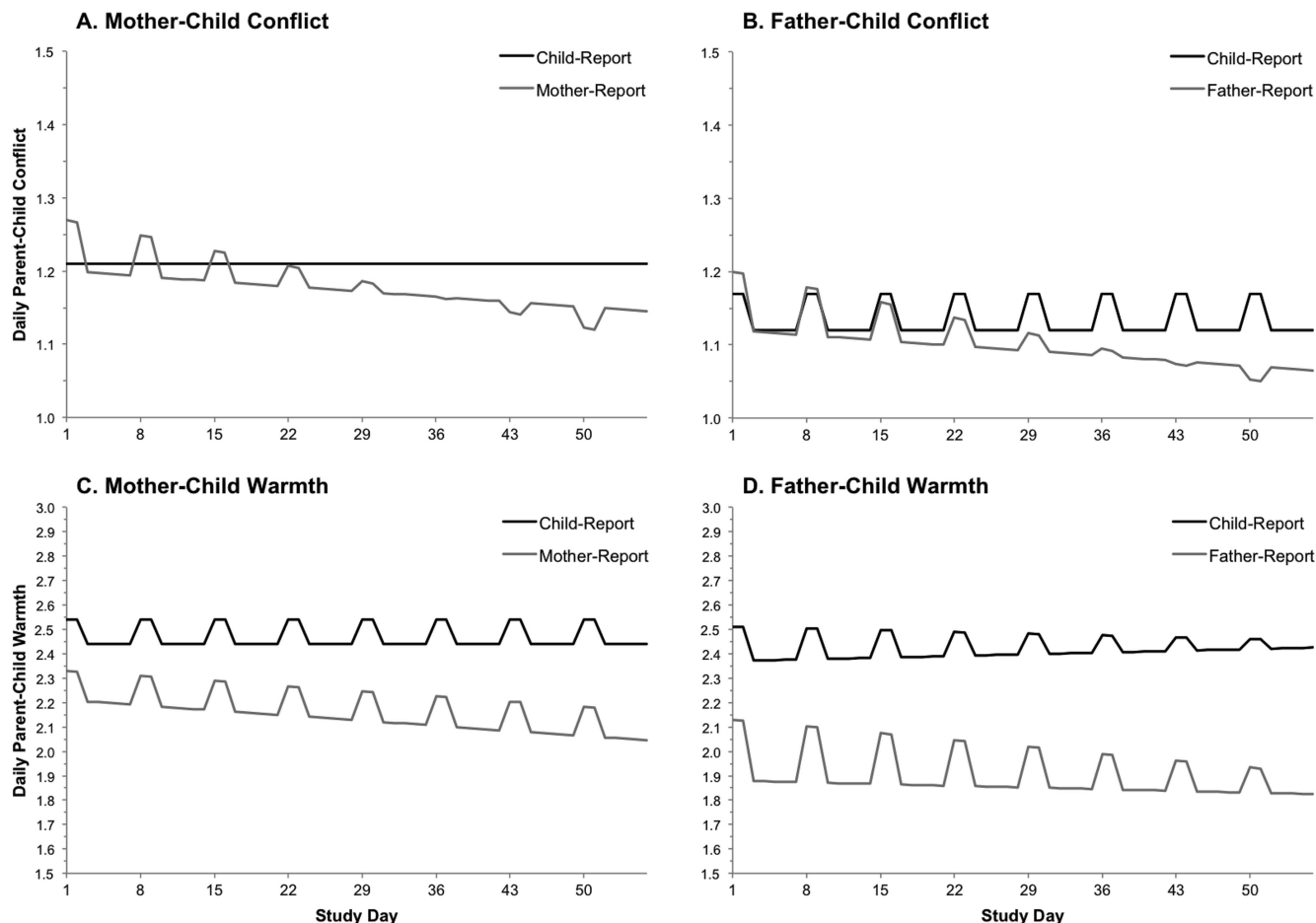


Figure 1. Predicted values based on multilevel models of parent–child conflict (top) and warmth (bottom) are depicted as a function of time in the study and weekdays versus weekend days. Markers on the x-axes represent the start of each new study week (always a Saturday). Weekend ratings appear as “bumps” in the trajectories.

warmth on days they reported more negative mood, and the strength of these associations remained stable over time.

In sum, weakening agreement between child and mother ratings of conflict, together with strengthening associations between child reports of negative mood and same-diary ratings of conflict, suggest potential fatigue-related declines in response accuracy among child participants. However, we observed no evidence of temporal changes in accuracy for daily reports of warmth.

Discussion

Despite the burden placed on parents and children by an 8-week daily diary study, participants showed compliance rates upward of 90%, comparable with previous electronic diary studies with adults across shorter time periods (Hufford, 2007; Shiffman et al., 2008). Measurement reactivity—conceptualized as systematic change in diary ratings as a function of time in the study—was generally minimal, less evident in child compared with parent reports, and in several cases, opposite of the directions predicted based on the self-monitoring literature. Although overall compliance was high, we observed some evidence consistent with fatigue, particularly among children.

Although effects were small, these results highlight areas for caution and point to design improvements that can maximize the benefits of diary research with families.

Measurement Reactivity

We found that child diary reports showed flat linear growth trajectories and, thus, presumably no measurement reactivity. On the other hand, both mother and father reports of parent–child conflict and warmth decreased across the study. Although benefits associated with mindful awareness and tracking of behavior (self-monitoring) might account for declines in parent reports of conflict, this cannot explain why ratings of warmth also decreased over time. Another interpretation is suggested by one of the only other studies to have examined reactivity in the family context (Merrilees et al., 2008): *response shift* (Schwartz et al., 2004) in which husbands recalibrated their standards for judging marital interactions over repeated assessments.

Specifically, parents began their diaries reporting relatively high levels of warmth and may have, over time, increased their thresholds for judging whether items such as “Today I praised my child” justified a response of “a lot” (vs. “some” or “not at all”) in order

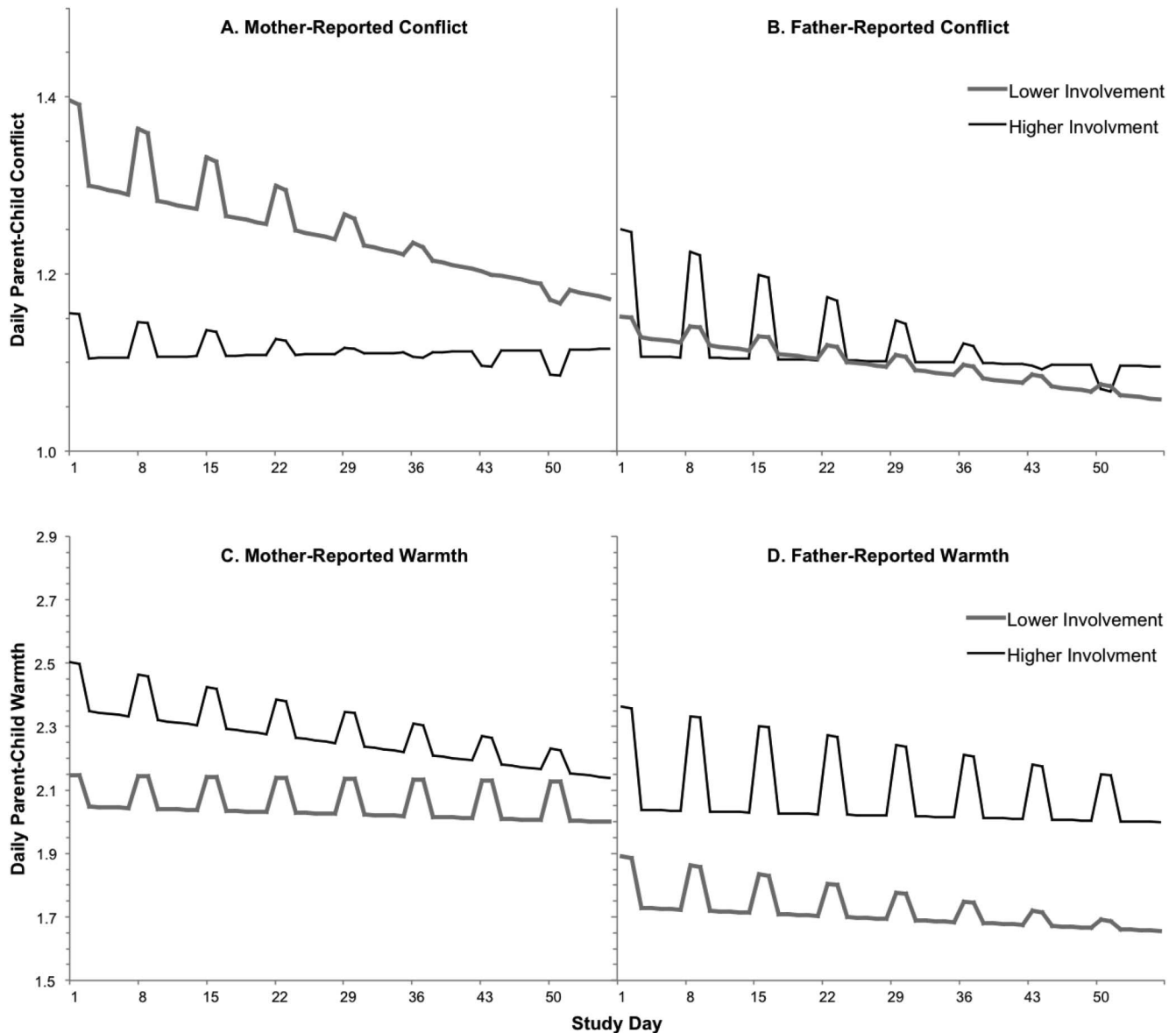


Figure 2. Model-implied trajectories of change in daily reports of parent-child conflict (top) and warmth (bottom) are depicted for prototypical mothers and fathers reporting higher (+1 *SD*) versus lower (-1*SD*) levels of global parent-child positive involvement at baseline. Markers on the *x*-axes represent the start of each new study week (always a Saturday). Weekend ratings appear as “bumps” in the trajectories.

to avoid ceiling problems and more accurately capture especially positive days with their children. Response shift can also account for the unexpected pattern of moderated change for mothers, in which the sharpest declines in daily warmth were observed among mothers who scored the highest on baseline measures of positive involvement. These mothers also had the highest initial diary ratings of warmth, and were likely most susceptible to shifting their responses downward over time in order to better capture day-to-day fluctuations in warmth.

On the other hand, parent daily reports of conflict were generally quite low and decreased across the study, a pattern not well explained by response shift but in line with predictions based on

the self-monitoring literature. Patterns of moderation for mother daily reports of conflict were also in line with self-monitoring hypotheses. Specifically, those with the lowest baseline ratings of positive involvement showed the sharpest declines in daily reports of conflict across the study.

Moderation results for father daily reports of conflict are more complex. Counter to expectations and findings observed among mothers, fathers reporting higher levels of positive involvement at baseline showed the sharpest declines in weekend ratings of conflict across the study. This implies that fathers with overall higher quality relationships with their children benefited most from daily monitoring of parent-child conflict.

One explanation for this finding could be that more positively involved fathers play a larger role in parenting overall, including in discipline and punishment, and therefore have more opportunities to reduce conflict. Such between-family differences might be most evident on weekend days when, compared with weekdays, parents typically have more time to spend interacting with children. Indeed, follow-up analyses revealed that more involved fathers reported spending significantly more time with their children on a daily basis compared with less involved fathers, and that parents in general spent more time with their children on weekend days compared with on weekdays.

It seems reasonable that behavioral shifts associated with repeated sampling would be most evident on weekend days when parents and children have more opportunities for interaction. Indeed, we found relatively consistent weekend effects, in which reports of parent-child conflict and warmth were higher on weekends than on weekdays, suggesting a natural rhythm facilitated by increased time spent together (Almeida et al., 2001). Interestingly, weekend and weekday levels of father-child warmth became more similar over time according to both father and child reports. Thus, repeated measurement may lead to greater consistency in father-child interactions. Alternatively, simultaneous participation in the study could have prompted increased warmth between fathers and children during the first few days of diary collection (which were always weekend days), with this effect diminishing across time in the study. Given the relatively small number of fathers in our sample (39 fathers vs. 47 mothers and children), further research is needed to confirm these patterns of change and explore potential mechanisms.

Further work is also needed to reconcile why temporal trends were evident in parent but not child daily reports. Perhaps parents are more sensitive to subtle shifts in daily interactions. Differences in parent versus child diary scale composition, particularly in our assessment of parent-child conflict, could be another contributing factor. We intentionally included fewer items on child diaries to reduce burden. As a result, parent-report scales were more comprehensive and were able to capture changes in a wider range of parent-child interactions. As we discuss shortly, fatigue may also help account for differences in parent versus child change trajectories.

Participant Burden and Fatigue

Our compliance rates are notable given that little has been published on the feasibility of extended diary studies with families, particularly those that incorporate child reports. Children between the ages of 8 and 13 completed an average of 53 out of the 56 expected diaries, 98% of which were verified by electronic stamp as having been done on the evening due or before 9:00 a.m. the next morning. This extends prior reports of good compliance with 2-week paper diary protocols among high school students (Chung et al., 2009; Kiang & Buchanan, 2014), to electronic diaries, in a younger sample, and across a longer diary period (with a similar weekly rate of compensation). Results also demonstrate the feasibility of collecting diaries from multiple family members, as high compliance was achieved among mothers, fathers, and children.

Although overall compliance was excellent, the predicted probability of diary completion was lower on weekend days compared

with on weekdays during the initial study week. One explanation is that participation began on a Saturday and, therefore, families who forgot to start their diaries on the correct day could have contributed to an initially lower weekend compliance rate. Although the probability of diary completion on weekends remained stable over time, we observed significant within-person declines in the odds of weekday diary completion. This may reflect a general drop in compliance over time, or accumulating effects of burden and fatigue due to competing work or school obligations on weekdays. Even during the last week of the study, however, the probability of weekday diary completion remained quite high at 95%. Anecdotally, during feedback interviews, children and parents described only minor burden or inconvenience associated with the diaries (Robles et al., 2013).

The goal of our larger study—to examine how daily family stress impacts susceptibility to upper respiratory infection—necessitated a lengthy diary period. Although compliance rates should be encouraging to researchers wishing to study similarly infrequent events, other findings suggest areas for caution. Even when diaries are completed as instructed, fatigue may lead to less careful or accurate responding over time. Our design involving multiple family members allowed us to address this issue by testing for changes in agreement between parent and child diary reports across the study. Although differences in parent versus child growth trajectories were evident on average, within-family agreement remained stable over time with only one exception: Same-day child and mother reports of conflict became less strongly associated across the study. Independently, this finding could be explained by factors other than fatigue, such as differential sensitivity of the parent versus child diary in detecting true changes in conflict over time.

However, together with weakening agreement between mother and child reports of conflict, results of bivariate models appear to implicate fatigue on the part of child participants. Specifically, children's reports of negative mood more strongly predicted their weekend ratings of conflict with both parents across time in the study. Even though weekend compliance rates remained stable, these results suggest that fatigue might have led to less careful responding over time, with mood-congruent retrieval increasingly influencing child diary reports.

That children reported more conflict with parents on days they reported more negative mood is in line with previous diary research (Flook, 2011; Kiang & Buchanan, 2014); however, absent measurement-related bias, we would not expect such associations to strengthen over time. This highlights the importance of testing for measurement effects not only in individual variables but also in bivariate associations, as children's daily reports of negative mood and parent-child conflict showed no evidence of temporal change independently. Without statistically controlling temporal trends in bivariate models, within-person associations with child mood may be inflated.

It should be noted, however, that reports of mood and conflict were made concurrently each day, so we cannot infer causation in either direction. It is possible that, by repeatedly calling attention to parent-child interactions, diary completion led children to feel worse on days characterized by higher conflict. Studies incorporating multiple diary reports across the day will be able to further elucidate these associations, but at the risk of increased burden.

Limitations and Future Directions

Our findings are qualified by several limitations, which suggest directions for future research. A number of study procedures may have contributed to burden and fatigue beyond diary completion, including life stress interviews, sleep diaries, and salivary cortisol samples. Given the substantial commitment required from multiple family members across 8 weeks, families willing to volunteer for studies like ours may differ from the larger population.

Our data came from a relatively small sample of families, which limits power to detect between-family effects; thus, caution is warranted in interpreting moderation results pending replication. Additional research will be particularly helpful in clarifying why baseline ratings of positive involvement, but not global perceptions of parent-child conflict, moderated change in parent diary ratings. Studies with larger samples will provide better power to examine a range of additional individual- and family-level moderators of reactivity and fatigue. For example, results may differ among older teens due to differences in cognitive and emotional development, as well as changes in the parent-child relationship across the adolescent years (e.g., declines in closeness and communication; increases in conflict; divergence between youth and parent perceptions of their relationship; Laursen & Collins, 2009). Child gender and pubertal development, parent marital status, and personality traits like neuroticism (Conner & Reid, 2012) may also help predict susceptibility to measurement-related change in diary studies with families.

Given the number of diaries collected from each family member, our study was adequately powered to detect even small within-person trends. Indeed, according to pseudo- R^2 statistics, the temporal changes we observed were generally small according to Cohen's (1988) guidelines. The average predicted decline in parents' weekend reports of conflict and warmth across the entire 8 weeks ranged between .15 and .20 on a 3-point response scale, with even smaller decreases across weekdays (.05 to .15). Significance tests made no corrections to alpha levels despite conducting multiple analyses. Thus, it is possible some significant results were obtained due to chance (Type I error); however, given the similarity of linear trends across mother and father reports, we suspect that temporal changes reflect more than statistical error.

Further investigation is needed to clarify how the observed temporal trends should be interpreted. Combining diary methods with pre- and postassessments, while experimentally manipulating sampling frequency between families, could help narrow the possibilities. Technological advancements have also allowed for more objective means of sampling experience, such as through video cameras placed in the home (Ochs & Kremer-Sadlik, 2013) or small devices known as Electronically Activated Recorders (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001), designed to unobtrusively record ambient sounds in participants' environment as they go about their day. These methods could be combined with diaries to assess whether measurement-related changes in self-reports are reflected in actual behaviors.

Further study is also needed to better understand response shift, and particularly the processes whereby participants conceptualize the meaning of diary items and calibrate their standards for responding. This information will be valuable in designing future measures, and could also point to confounding influences if response shift varies systematically with con-

structs under investigation. Extensive pilot testing and participant training is critical in minimizing this possibility and ensuring unambiguous interpretation of diary prompts and response options.

Descriptive studies like ours only scratch the surface of addressing methodological concerns in diary research. Although we focused on changes in mean levels of constructs and the strength of same-day associations over time, both measurement reactivity and fatigue can also be indicated by changes in variability (Moskowitz, Russell, Sadikaj, & Sutton, 2009). Moreover, measurement reactivity is not necessarily linear and may be characterized by sudden jumps in scores after the first few days of reporting, or at other meaningful points throughout the study.

Beyond those already discussed, the current findings have several implications for diary research on parent-child interactions. If daily diaries are used to assess the effectiveness of parenting interventions, for instance, the measurement-related changes identified here could introduce noise and obscure treatment effects. If repeated assessment can alter natural, everyday family interactions, then this has clear implications for ecological validity. On the other hand, rather than signaling measurement-induced shifts in behavior, the temporal trends observed here could reflect increasingly accurate representations of natural family life. In such case, diary studies spanning longer than one or two weeks might prove more ecologically valid. Any improvements in validity, however, would have to be weighed against the potential pitfalls of increased participant burden, especially in studies involving children.

In sum, multiple family members, including a child between the ages of 8 and 13, showed remarkably good compliance with an 8-week diary protocol. When present, parent reactivity effects were generally small and not always in directions predicted by self-monitoring literature. Some evidence consistent with child fatigue highlighted the importance of examining temporal changes in bivariate associations and agreement between reporters. We have suggested a number of avenues for future research that could build on and further elucidate the findings reported here, thus contributing to our growing understanding of the methodological challenges, as well as benefits, of diary designs in developmental research with families.

References

- Almeida, D. M., Wethington, E., & McDonald, D. A. (2001). Daily variation in paternal engagement and negative mood: Implications for emotionally supportive and conflictual interactions. *Journal of Marriage and Family, 63*, 417-429. <http://dx.doi.org/10.1111/j.1741-3737.2001.00417.x>
- Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108-123). London, UK: Guilford Press.
- Broderick, J. E., Schwartz, J. E., Shiffman, S., Hufford, M. R., & Stone, A. A. (2003). Signaling does not adequately improve diary compliance. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine, 26*, 139-148.
- Chung, G. H., Flook, L., & Fuligni, A. J. (2009). Daily family conflict and emotional distress among adolescents from Latin American, Asian, and European backgrounds. *Developmental Psychology, 45*, 1406-1415. <http://dx.doi.org/10.1037/a0014163>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Conner, T. S., & Reid, K. A. (2012). Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science*, 3, 315–323. <http://dx.doi.org/10.1177/1948550611419677>
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, 32, 917–929. <http://dx.doi.org/10.1177/0146167206287721>
- Doumas, D. M., Margolin, G., & John, R. S. (2003). The relationship between daily marital interaction, work, and health-promoting behaviors in dual-earner couples: An extension of the work-family spillover model. *Journal of Family Issues*, 24, 3–20. <http://dx.doi.org/10.1177/0192513X02238518>
- Elkins, I. J., McGue, M., & Iacono, W. G. (1997). Genetic and environmental influences on parent-son relationships: Evidence for increasing genetic influence during adolescence. *Developmental Psychology*, 33, 351–363. <http://dx.doi.org/10.1037/0012-1649.33.2.351>
- Flook, L. (2011). Gender differences in adolescents' daily interpersonal events and well-being. *Child Development*, 82, 454–461. <http://dx.doi.org/10.1111/j.1467-8624.2010.01521.x>
- Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 54–75). New York, NY: Oxford University Press.
- Johnson, E. I., Grondin, O., Barrault, M., Faytout, M., Helbig, S., Husky, M., . . . Swendsen, J. (2009). Computerized ambulatory monitoring in psychiatry: A multi-site collaborative study of acceptability, compliance, and reactivity. *International Journal of Methods in Psychiatric Research*, 18, 48–57. <http://dx.doi.org/10.1002/mpr.276>
- Kiang, L., & Buchanan, C. M. (2014). Daily stress and emotional well-being among Asian American adolescents: Same-day, lagged, and chronic associations. *Developmental Psychology*, 50, 611–621. <http://dx.doi.org/10.1037/a0033645>
- Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (1999). Emotion and memory: Implications for self-report. In A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 81–100). Mahwah, NJ: Lawrence Erlbaum.
- Kirkpatrick, S. I., Midthune, D., Dodd, K. W., Potischman, N., Subar, A. F., & Thompson, F. E. (2012). Reactivity and its association with body mass index across days on food checklists. *Journal of the Academy of Nutrition and Dietetics*, 112, 110–118. <http://dx.doi.org/10.1016/j.jada.2011.10.004>
- Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, 11, 415–425. <http://dx.doi.org/10.1037/1040-3590.11.4.415>
- Laursen, B., & Collins, W. A. (2009). Parent—Child relationships during adolescence. In R. M. Lerner & L. D. Steinberg (Eds.), *Handbook of adolescent psychology: Vol. 2. Contextual influences on adolescent development* (3rd ed., pp. 3–42). New York, NY: Wiley.
- Lehman, B. J., & Repetti, R. L. (2007). Bad days don't end when the school bell rings: The lingering effects of negative school events on children's mood, self-esteem, and perceptions of parent-child interaction. *Social Development*, 16, 596–618. <http://dx.doi.org/10.1111/j.1467-9507.2007.00398.x>
- Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology*, 17, 48–52. <http://dx.doi.org/10.1037/0278-6133.17.1.48>
- Maas, J., Hietbrink, L., Rinck, M., & Keijsers, G. P. J. (2013). Changing automatic behavior through self-monitoring: Does overt change also imply implicit change? *Journal of Behavior Therapy and Experimental Psychiatry*, 44, 279–284. <http://dx.doi.org/10.1016/j.jbtep.2012.12.002>
- Magneberg, R. (1998). The effect of electronic surveillance on alcohol misuse and everyday behavior. *Addictive Behaviors*, 23, 281–301. [http://dx.doi.org/10.1016/S0306-4603\(97\)00067-1](http://dx.doi.org/10.1016/S0306-4603(97)00067-1)
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments & Computers*, 33, 517–523. <http://dx.doi.org/10.3758/BF03195410>
- Merrilees, C. E., Goeke-Morey, M., & Cummings, E. M. (2008). Do event-contingent diaries about marital conflict change marital interactions? *Behaviour Research and Therapy*, 46, 253–262. <http://dx.doi.org/10.1016/j.brat.2007.11.007>
- Moskowitz, D. S., Russell, J. J., Sadikaj, G., & Sutton, R. (2009). Measuring people intensively. *Canadian Psychology*, 50, 131–140. <http://dx.doi.org/10.1037/a0016625>
- Nezlek, J. B. (2003). Using multilevel random coefficient modeling to analyze social interaction diary data. *Journal of Social and Personal Relationships*, 20, 437–469. <http://dx.doi.org/10.1177/02654075030204002>
- Nishina, A. (2012). Adolescent daily and general maladjustment: Is there reactivity to daily repeated measures methodologies? *Child Development*, 83, 405–412.
- Ochs, E., & Kremer-Sadlik, T. (2013). *Fast-forward family: Home, work, and relationships in middle-class America*. Berkeley, CA: University of California Press.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. <http://dx.doi.org/10.3102/10769986031004437>
- Repetti, R. L. (1996). The effects of perceived daily social and academic failure experiences on school-age children's subsequent interactions with parents. *Child Development*, 67, 1467–1482. <http://dx.doi.org/10.2307/1131712>
- Repetti, R. L., Reynolds, B. M., & Sears, M. S. (2015). Families under the microscope: Repeated sampling of perceptions, experiences, biology and behavior. *Journal of Marriage and Family*, 77, 126–146. <http://dx.doi.org/10.1111/jomf.12143>
- Repetti, R. L., Robles, T. F., & Reynolds, B. (2011). Allostatic processes in the family. *Development and Psychopathology*, 23, 921–938. <http://dx.doi.org/10.1017/S095457941100040X>
- Repetti, R. L., Robles, T. F., & Reynolds, B. M. (2014). Biological and psychological processes linking chronic family stress to substance abuse and obesity. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (2nd ed., pp. 398–418). New York, NY: Guilford Press.
- Robles, T. F., Reynolds, B. M., Repetti, R. L., & Chung, P. J. (2013). Using daily diaries to study family settings, emotions, and health in everyday life. *Journal of Social and Personal Relationships*, 30, 179–188. <http://dx.doi.org/10.1177/0265407512457102>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. <http://dx.doi.org/10.1037/1082-989X.7.2.147>
- Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology & Health*, 19, 51–69. <http://dx.doi.org/10.1080/0887044031000118456>
- Sears, M. S., Repetti, R. L., Reynolds, B. M., & Sperling, J. B. (2014). A naturalistic observational study of children's expressions of anger in the family context. *Emotion*, 14, 272–283. <http://dx.doi.org/10.1037/a0034753>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. London, UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195152968.001.0001>

Surra, C. A., Curran, M. A., & Williams, K. (2009). Effects of participation in a longitudinal study of dating. *Personal Relationships, 16*, 1–21. <http://dx.doi.org/10.1111/j.1475-6811.2009.01207.x>

Telzer, E. H., & Fuligni, A. J. (2013). Positive daily family interactions eliminate gender differences in internalizing symptoms among adolescents. *Journal of Youth and Adolescence, 42*, 1498–1511. <http://dx.doi.org/10.1007/s10964-013-9964-y>

Timmons, A. C., & Margolin, G. (2015). Family conflict, mood, and adolescents' daily school problems: Moderating roles of internalizing

and externalizing symptoms. *Child Development, 86*, 241–258. <http://dx.doi.org/10.1111/cdev.12300>

Received October 22, 2014

Revision received September 22, 2015

Accepted October 20, 2015 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit <http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx>.