

## Changes in the Relation of Self-Efficacy Beliefs and Behaviors Across Development

Pamela E. Davis-Kean, L. Rowell Huesmann,  
and Justin Jager  
*University of Michigan*

W. Andrew Collins  
*University of Minnesota*

John E. Bates  
*Indiana University*

Jennifer E. Lansford  
*Duke University*

Many social science theories that examine the connection between beliefs and behaviors assume that belief constructs will predict behaviors similarly across development. Converging research implies that this assumption may not be tenable across all ages or all belief constructs. Thus, to test this implication, the relation between behavior and beliefs about the self was examined in 2 independent data sets with 2 different constructs: aggression and achievement. The respondents were 6–18 years of age and predominately Caucasian. Results using quasi-simplex structural equation models suggest that self-beliefs become more strongly related to behavior as children grow older independent of the reliability of the measures used. Possible limitations in the use of self-report methodology with young children are discussed.

Children entering school are confronted with two important tasks: to achieve and to get along cooperatively with others. Because of the importance of these two tasks for positive development, many researchers have focused their research on trying to understand what might be the major influences on these outcomes. An important tenet of this work is that a link exists between beliefs about ability to perform a behavior (self-efficacy) and actually performing the behavior (e.g., Bandura, 1986; Crick & Dodge, 1994; Huesmann, 1988; Kohlberg, 1969; Wigfield & Eccles, 1992). The theoretical models based on these tenets are used to explain children's behaviors by the various belief domains that may be specific to the behavior. For example, when trying to understand achievement outcomes, children's beliefs about their ability to perform are often assessed and used to predict behavior (Bandura, 1986; Wigfield & Eccles, 1992). A similar method is used in research in which

children's beliefs about appropriate behavior with peers are assessed in connection with their aggressive behaviors (Crick & Dodge, 1994; Huesmann & Guerra, 1997).

The models guiding this research imply that children's beliefs and expectations about their behavior will dictate or predict their behaviors and that this prediction is similar across development. Younger children, for example, would show similar relations between their beliefs about achievement and actual achievement as would older children. An assumption from these theories, then, is that a primary intervention for changing children's behaviors in achievement and peer relations is to change their beliefs about important behaviors and that the age of the child would not be a factor in this intervention. Even though these models linking beliefs and behaviors have been well examined across the years, the general focus has been on cross-sectional relations or short-term longitudinal studies with very limited age ranges (e.g., middle childhood or adolescence). This limitation primarily reflects the lack of longitudinal data that spanned from early schooling (e.g., first/second grade) to later schooling (high school). Thus, little research has examined the specific relation between self-beliefs and their associated behaviors across time.

---

This research was supported by the Center for the Analysis of Pathways from Childhood to Adulthood, which is funded by National Science Foundation Grant 0322356. We thank the following people for contributions of their data sets to this study: Kenneth A. Dodge and Gregory S. Pettit (Child Development Project), and Jacque Eccles, Alan Wigfield, Rene Harold, and Phyllis Blumenfeld (Childhood and Beyond Study). We are also very grateful for the statistical assistance of Chris Mutch, Jeremy Welland, Holly Sexton, and Jackson Goodnight.

Correspondence concerning this article should be addressed to Pamela E. Davis-Kean, 426 Thompson Street, University of Michigan, Ann Arbor, MI 48106-1248. Electronic mail may be sent to pdakean@umich.edu.

© 2008, Copyright the Author(s)  
Journal Compilation © 2008, Society for Research in Child Development, Inc.  
All rights reserved. 0009-3920/2008/7905-0004

A primary goal of this article is to examine the relation between self-efficacy beliefs and behaviors across time to determine the stability of this relation. A question of particular interest is whether this relation remains stable across time or changes as children age. A related question is whether, if the belief-behavior relation changes, the change is related to age. A second goal of this article is to examine these relations for two important developmental outcomes: achievement and negative peer relations (aggression). This type of research is important for understanding whether self-referent beliefs about behaviors are similarly predictive across time. If self-referent beliefs do not relate to behaviors similarly across time, then it is important to understand how these relations may change so that interventions may be appropriately targeted for the age of the child.

#### *The Development of Beliefs and Behaviors*

Considerable research has suggested that younger children have difficulties with certain tasks and concepts (e.g., future expectations, predictions of others' behaviors, conservation tasks) but that these difficulties are less apparent at older ages (Alvarez, Ruble, & Bolger, 2001; Case, 1991; Harter, 1996; Parsons & Ruble, 1977; Ruble, Parsons, & Ross, 1976). Piaget (1963/2001; Siegler & Alibali, 2004) attributed these changes to a shift from the preoperational stage of reasoning to a concrete operational stage, which involves more abstract thinking and use of symbolic concepts beginning at about age 7. Although previous cognitive development research focused on the idea that children develop through discrete, qualitative stages, researchers currently emphasize that children steadily become better and more efficient information processors as the brain develops a greater capacity for more complicated procedures (e.g., parallel processing; Siegler & Alibali, 2004). Evidence of mental capabilities that children did not possess at earlier ages continues well into adolescence (Keating, 2004).

This revised view does not mean, however, that development is linear. For example, some dramatic changes appear to occur between the ages of 5 and 8, making this an important focus for research (Siegler & Alibali, 2004). Indeed, 8-year-olds have become the most commonly used age group in studies of children's ability to reason about themselves and others (Alvarez et al., 2001; Harter, 1996). Researchers have referred to this change as the "5-7 year shift" (Sameroff & Haith, 1986; White, 1970). White (1996), for example, referred to the time period after the age of 7 as entering the "age of reason." Historically, ages 5-7 have been regarded as an important turning point distinguishing child-

hood from early adulthood because of the changes in the mental ability of the child. However, there is also research to suggest that adolescence may be a key time of change as children's cognitive development continues to mature and shifts to reasoned action and reflection (Keating, 2004). It is important when examining children's beliefs to know whether or not significant shifts occur in children's ability to understand and then report on their beliefs across time in order to understand when these beliefs might influence behavior.

#### *Development of Belief-Behavior Relations*

Though few direct tests exist of the developmental change in belief-behavior relations, differences in belief-behavior connections before and after 8 years of age have been reported frequently. Findings from developmental research suggest that younger children's beliefs and behaviors may be only negligibly related before the age of 7 (Parsons & Ruble, 1977; Spencer & Bornholt, 2003) and that beliefs are unreliable predictors of future behaviors until after the age of 8 (Huesmann & Guerra, 1997). This phenomenon has been documented in research on such diverse domains as self-esteem (Davis-Kean & Sandler, 2001), expectancies regarding success and failure (Parsons & Ruble, 1977), beliefs about the acceptability of aggression (Huesmann & Guerra, 1997), and social information processing in relation to externalizing behavior (Lansford et al., 2006). These arrays of findings give some indication that the relations between self-efficacy beliefs and behaviors may vary across development.

Further support comes from research on expectancies for success and failure. Parsons and Ruble (1977), for example, showed that older children incorporated information regarding their outcomes quickly and changed their predictions about future successes or failures accordingly, whereas children younger than age 8 typically did not. This is consistent with developmental work by Harter (1999) suggesting that possible cognitive limitations to young children's ability to answer these types of questions (self-reflective) at earlier ages.

These findings, though suggestive, do not constitute direct evidence on the question of whether stability or change occurs in the relation between self-efficacy beliefs and behaviors across development. In order to test for stability, one would need to examine the synchronous correlations between self-efficacy behavioral beliefs and their associated behaviors and see if they differ across time. If they do not significantly differ across time, then these relations are stable. Conversely, if the synchronous correlations

do significantly differ across time, then these relations are changing and it will be important to specify at what age these relations may be changing.

The current study is intended to describe and test whether or not the relations between self-efficacy beliefs and behaviors stay stable or change across time. We examine these relations across time by using information from multiple data sets and multiple belief domains. If the relations change, are the relations weaker in childhood than in adolescence, as previous research appears to suggest or do the associations differ from one age period to another across early childhood, middle childhood, and adolescence? An important consideration in examining these changes is whether reliability of the measures is a factor in the change, such that younger children are less reliable in their answers, thus attenuating the correlations between beliefs and behaviors (Davis-Kean & Sandler, 2001; Trzesniewski, Donnellan, & Robins, 2003).

We use two different longitudinal data sets with observations collected both in childhood and in adolescence. In these data sets, we have information on two different behaviors (i.e., academic achievement and aggressiveness) and self-efficacy beliefs regarding these behaviors (e.g., self-concept of ability and self-efficacy of aggression). Thus, a unique feature of this study is that we will test the generalizability of change in associations between self-efficacy beliefs and behaviors both through analyzing two independent data sets and by using two different important childhood outcomes.

## Method

The data sets for the study come from a collaboration of longitudinal researchers who have made their data sets available to answer and replicate developmental questions. The data sets selected for the analyses met two criteria. First, both had been guided by theories (e.g., information processing theories, expectancy-value theory) positing similar self-efficacy belief–behavior relations across development. Thus, the researchers had collected data that measured beliefs and behaviors on multiple occasions across childhood and adolescence. Second, the self-efficacy beliefs and behaviors were assessed in a similar manner across time. The two data sets that met these criteria and their relevant methodology are summarized below.

### *Data Sets and Measures*

*Childhood and Beyond (CAB).* Three cohorts of children from 28 schools located in four southeastern

Michigan school districts provided data for the current analysis (Eccles, Wigfield, Harold, & Blumenfeld, 1993). Academic belief (math self-concept of ability) and behavior (math grades) data were collected starting in the fall and spring of the 1988–1989 school year, respectively, from children in Grade 1 ( $n = 317$ ), Grade 2 ( $n = 330$ ), and Grade 4 ( $n = 423$ ) each year for 3 years. In a follow-up study starting in 1994, data were collected from these cohorts each year for 3 years. Thus, the Grade 1 cohort has data from six periods (Grades 1, 2, 3, 7, 8, and 9), the Grade 2 cohort has data from six periods (Grades 2, 3, 4, 8, 9, and 10), and the Grade 4 cohort has data from six periods (Grades 4, 5, 6, 10, 11, and 12). Ninety-three percent of participants in the sample are Caucasian, 50% are female, and all generally were of relatively high socioeconomic status (SES).

Self-concept of math ability (SCMA) and math grades served as the primary belief and behavior variables of interest, respectively. SCMA is a composite variable that includes five items: “How good at math are you?” “If you had to list all the students from best to worst in math where are you?” “Compared to other subjects how good are you at math?” “How well do you expect to do in math this year?” and “How good would you be at learning something new in math?” Responses were measured on a Likert-type scale ranging from 1 to 7 (e.g., 1 = *one of the worst*, 7 = *one of the best*) and collected in the spring of the child’s academic year. SCMA Cronbach’s alphas were .76, .79, .85, .90, .92, and .93 for Waves 1 through 6, respectively.

Math grades were obtained from school records. In the first three waves, math grades correspond to the child’s math performance during the spring only; in the latter three time periods, the child’s final math grade for the academic year was used. Grades were coded on a scale of 1–16 where 1 represented an F– and 16 an A+. The grades ranged between 1 and 16 for the Grade 1 cohort, between 3 and 16 for the Grade 2 cohort, and between 1 and 16 for the Grade 4 cohort.

*Child Development Project (CDP).* The families in the CDP (Dodge, Bates, & Pettit, 1990) were recruited when the children entered kindergarten in 1987 or 1988 at three sites: Knoxville and Nashville, TN, and Bloomington, IN. The combined sample consisted of 585 families at the first assessment. Males comprised 52% of the sample. Eighty-one percent of the sample were Caucasian, 17% were African American, and 2% were from other ethnic groups. Follow-up assessments used in the present study were conducted in Grades 1–3, 8, and 11, giving a total of six waves of data. Compared to the original sample, the families who provided data in Grade 11 (78% of the original sample) were of

slightly higher SES, but participants and nonparticipants did not differ by race, single-parent status, or mothers' reports of children's externalizing behaviors in kindergarten.

In Grades kindergarten through 3, children were presented with a series of four cartoon pictures and brief verbal descriptions of the cartoon events. In each cartoon, children were told to imagine being the protagonist. After hearing the description of each event, children were presented with alternative strategies (competent, aggressive, and inept) for dealing with the situation. Efficacy for aggression was assessed by asking children to rate how easy it would be for them to act aggressively in each situation (1 = *HARD*, 2 = *hard*, 3 = *easy*, 4 = *EASY*). Efficacy for aggression was scored as the average of this item following the aggressive response across the four vignettes. Cronbach's alphas for this measure in Grades 1–3 were .64, .59, and .55, respectively. In Grades 8 and 11, participants were shown six video vignettes that started with a social interaction and culminated in an ambiguous provocation by peers or adults directed toward a protagonist adolescent. Participants were asked to imagine being the protagonist. Three vignettes depicted provocations that were relevant to both boys and girls and were presented to all participants; the remaining three vignettes were sex specific. Each provocation segment was followed by two segments that presented an aggressive and a nonaggressive response to the provocation, respectively. After each of the responses presented on the video, efficacy for aggression was assessed by asking adolescents how easy it would be for them to act aggressively in each situation (1 = *very hard*, 5 = *very easy*). A composite variable was created in Grade 8 and Grade 11 by averaging this item from the aggressive responses across the six vignettes. Cronbach's alphas for these measures were .82 and .86, respectively.

Mothers completed the Child Behavior Checklist (CBCL; Achenbach, 1991) annually when their children were in kindergarten through grade 11. Mothers reported whether each item was 0 = *not true*, 1 = *somewhat or sometimes true*, or 2 = *very or often true* of the child. The 20 items in the aggression subscale were summed to create a separate scale for each year. Cronbach's alphas were high and stable (.87, .87, .87, .89, .89, .90 for Waves 1 through 6, respectively).

#### *Attrition and Missing Data Plan*

The problem of missing data is inherent in the use of longitudinal data sets. Each of the data sets used in this study had various amounts of missing data—ranging from a low amount of missing (CDP ~ 5%) to

a rather high amount of missing (CAB ~ 40%). In order to minimize bias due to attrition, the CAB and CDP missing data were multiply imputed using the NORM program (Schafer, 1999). Compared to other missing data procedures, multiple imputation (MI) is superior when it comes to minimizing bias due to attrition (Graham, Hofer, & Piccinin, 1994; Little & Rubin, 1987; Schafer & Olsen, 1998). In the same regard, MI is also far superior to simple casewise deletion (Little & Rubin, 1987; Schafer & Olsen, 1998). Moreover, not only are the parameter estimates more trustworthy when using MI (due to its accounting for attrition more effectively) but also standard errors associated with the parameter estimates are not artificially decreased, and as such inferences are trustworthy as well (Graham et al., 1994; Schafer & Olsen, 1998).

One assumption of the MI procedure is that all variables included in the procedure are approximately normally distributed (Schafer & Olsen, 1998). Any data with violations of normality (e.g., aggressive behavior variables) were transformed prior to imputation. A second assumption of the MI procedure is that the data imputed are missing at random. When utilizing MI, the first step is to empirically identify as many variables as possible that are related to the missingness of the variables to be imputed. Once identified, these variables are included in the imputation process as "auxiliary variables," and their inclusion helps to ensure that the imputed data resulting from the MI procedure are unbiased. However, once the data are imputed, the auxiliary variables are dropped, and the imputed data are then analyzed separately from the auxiliary variables. Because the auxiliary variables are not included in the actual analyses, they are not equivalent to control variables.

Separately for CAB and CDP, the number of missing time points for each variable was used as a grouping variable, and a series of one-way analyses of variance were conducted to determine what auxiliary variables were significantly related to missingness. Several auxiliary variables were identified and incorporated into the MI model for both data sets, including SES indicators (i.e., income, education, and occupation), IQ measures (e.g., Slossan, Wechsler Intelligence Scale for Children–Revised), sex, race/ethnicity, and demographic location variables such as neighborhoods or school districts.

Due to the cohort-sequential design of CAB, there are many instances of planned missingness (e.g., a respondent is missing data due to the design of the study as opposed to nonresponse). In instances of planned missingness, data were not imputed. For example, due to the design of the study, no

respondent in the first cohort of CAB provided data at Grades 4, 5, 6, 10, 11, and 12. Thus, individuals in the first cohort of CAB did not have data imputed at these grades, and instead only had data imputed for the grades where data were actually collected (Grades 1, 2, 3, 7, 8, and 9). In a similar fashion, for the remaining two CAB cohorts, data were imputed only for the grades where data were actually collected. For each respondent, data were imputed at the individual item level (as opposed to the scale level). This was done for three reasons. First, when using MI and imputing scale data, it is standard practice to carry out those imputations at the item level (Schafer & Olsen, 1998). Second, when possible, we model the multiitem scales as latent constructs in our structural equation models, and therefore, our analyses demand item-level data. Finally, as is standard practice when utilizing MI (Graham et al., 1994; Schafer & Olsen, 1998), in instances where respondents had complete missing data for a given measure at a particular grade (whether based on a single-item or a multiitem scale), and that complete missingness was not due to planned missingness, data were imputed.

Among the 1,070 respondents included in the CAB analyses, the mean percentage missing across the 36 measures for which missing values were imputed (five observations for the math self-concept of ability scale at six time points and one observation of grades at six time points) was 40.1%. Based on this fraction, eight data sets were imputed to reach an efficiency level above .95 (Rubin, 1987). Among the 585 respondents included in the CDP analyses, the mean percentage missing across the 148 measures for which missing values were imputed (20 observations for the CBCL at six time points, 4 observations for the efficacy for aggression scale at the four initial time points, and 6 observations for the efficacy for aggression scale at the last two time points) was 5.3%. Based on this fraction, only three data sets needed to be imputed in order to reach an efficiency level above .95 (Rubin, 1987); however, because five imputed data sets is the standard minimum for number of imputed data sets (Schafer & Olsen, 1998), five data sets were imputed to reach an efficiency level well above .95.

### *Analysis Strategy*

Imputed means and estimates of their standard errors were calculated using PROC MIANALYZE within SAS (SAS Institute, 1999), which was developed to estimate standard errors of multiply imputed data. All analyses were conducted within Mplus (Muthen & Muthen, 1998–2006), using the imputation option developed for multiply imputed data.

In order to assess if the association between SCMA and math grades changes over time, we used the quasi-simplex measurement model presented in Figure 1. In order to assess if the relation between efficacy for aggression and aggressive behavior changes over time, we used the quasi-simplex measurement model presented in Figure 2. First, the models in Figures 1 and 2 are measurement models (as opposed to causal path models) because there are no causal paths in either model and, instead, only correlations are included. The inclusion of causal paths (say, autoregressive paths) would answer a different question—namely, “how does the relationship between changes in beliefs and changes in behaviors between time  $t$  and time  $t-1$  itself change over time?”

Second, the models in Figures 1 and 2 are both quasi-simplex models because in both models the behavior measure (i.e., math grade for the model in Figure 1 and aggressive behavior for the model in Figure 2) is based on a single indicator. Within a quasi-simplex model, in order to adjust for the reliability of each measure at each time point, errors are fixed based on the following equation:  $e = \sigma^2[1 - \alpha]$  (Joreskog, 1970). For the model in Figure 1, the measure for behavior (i.e., final math grade) is based on a single indicator; as such the interitem reliability between fall math grade and final math grade was used as a proxy measure for interitem reliability (allowing for the fact that grades generally represent a true score with some error). For each of the six points in time, a given point in time's interitem reliability between fall math grade and final math grade was applied. Across the six time periods, the alphas were .85, .85, .88, .95, .96, and .96, respectively. For the model in Figure 2, the measure for behavior (i.e., aggressive behavior) is based on the aggregate of 20 CBCL indicators for which the interitem reliability at each time point was applied. Across the six periods, the alphas were .87, .87, .87, .89, .89, and .90, respectively. For both of the latent behavior constructs (math grades and aggressive behavior), it is not possible to correlate the error variances of matching items across time because they are based on a single observed indicator (Marsh, 1993). However, for both the latent belief constructs (SCMA and efficacy for aggression), which were based on multiple observed indicators, the error variances of matching items were correlated across time.

In order to empirically assess whether or not the relation between behaviors and beliefs changes over time, we compared models where the synchronous covariances between beliefs and behaviors are constrained to be equal across time to models where the synchronous covariances are free to vary across time.

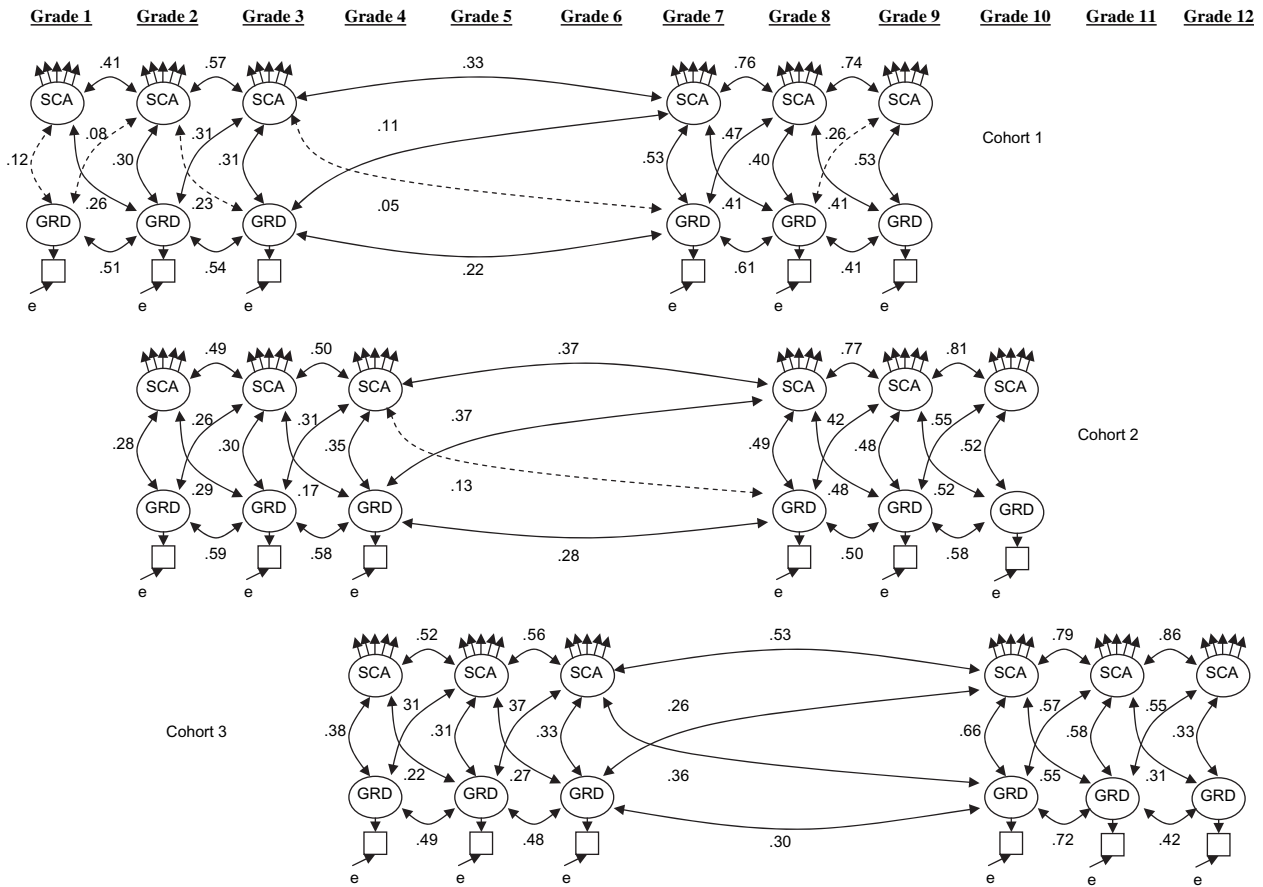


Figure 1. Estimates for quasi-simplex measurement model assessing relationship between math self-concept and math grades over time, by cohort, Childhood and Beyond Study.  
 Note. Important correlations are bolded for emphasis. Solid lines indicate significance at .05 level. SCA self-concept of ability; GRD short for math grade. For SCA, exogenous arrows represent indicators listed in Table 2. Fit indices:  $\chi^2(1,425) = 3,395.58, p < .001$ ; comparative fit index = .92; root mean square error of approximation = .06.

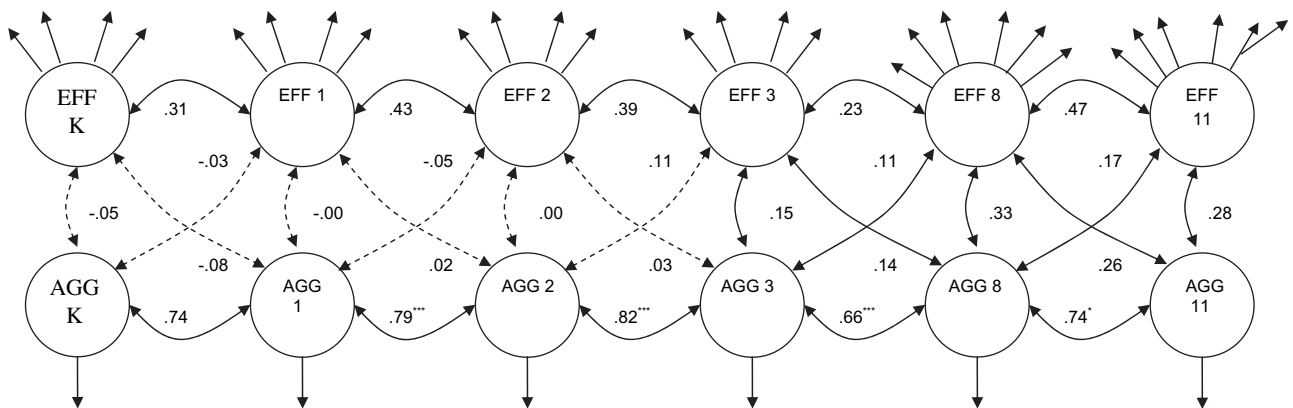


Figure 2. Estimates for quasi-simplex measurement model assessing relationship between aggression self-efficacy and aggression over time, Child Development Project.  
 Note. Important correlations are bolded for emphasis. Solid lines indicate significance at .05 level. EFF short for aggression self-efficacy; AGG short for aggression. For EFF, exogenous arrows represent indicators listed in Table 2. Fit indices:  $\chi^2(922) = 1,682.52, p < .001$ ; comparative fit index = .90; root mean square error of approximation = .05.

According to convention, model comparisons were based on changes in chi-square relative to changes in degrees of freedom (Kline, 1998). The chi-square value used in the model comparisons is the chi-square value provided in the Mplus output, which is the average chi-square value (Muthen, 2005) across the *m* data sets, where *m* is the number of imputed data sets.

**Results**

*Basic Descriptive Statistics*

The imputed means and standard deviations for all measures for CAB and CDP are presented in Table 1. The means and standard deviations are presented separately for each cohort. The factor loadings for the latent belief constructs, math self-concept of ability (CAB), and efficacy for aggression (CDP), and the latent behavior constructs, math grades, and aggressive behavior, were computed from the measurement models in Figures 1 and 2 and are listed in Table 2. Also listed in Table 2 are the residual variances for each observed variable included in the measurement models in Figures 1 and 2. Because there are eight imputed data sets for CAB and five imputed data sets for CDP, there are 13 covariance matrices in all. As such the covariance matrices are not presented here but are available from the authors upon request.

*Changes in Synchronous Correlations Between Math Self-Concept and Math Grades*

Before examining whether or not the association between math self-concept and math grades shown in Figure 1 changes over time, we tested for cohort and sex differences in the relations shown in the model. First, we tested for cohort differences. Specifically, a fully unconstrained model (Model 1, Table 3) where all covariances were free to vary across the three cohorts was compared to both a model where just the synchronous covariances were constrained to be equal across the three cohorts (Model 2, Table 3) and a model where all covariances were constrained to be equal across the three cohorts (Model 3, Table 3). The  $\chi^2_{diff}$  test showed that these models differed significantly from each other, and therefore, it was not tenable that these associations were same across the cohorts, Model 2,  $\chi^2_{diff}(12) = 60.37, p < .001$ , and Model 3,  $\chi^2_{diff}(52) = 197.19, p < .001$ . Next, we tested for sex differences. Again, we compared a fully unconstrained model (Model 4, Table 3) where all covariances were free to vary across sex to both a model where just the synchronous covariances were constrained to be equal across sex (Model 5, Table 3) and a model where all covariances were constrained to be equal across sex (Model 6, Table 3). The fit of Model 4 (the fully unconstrained model) did not fit the data any better than did Model 5,  $\chi^2_{diff}(6) = 7.96, p = .24$ , or

Table 1  
Overall Means for CAB and CDP, by Cohort

	N	Point of assessment					
		Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
<b>CAB</b>							
Cohort 1	317						
Math SCA		5.62 (1.03)	5.53 (0.98)	5.54 (1.00)	4.94 (1.25)	4.69 (1.28)	4.60 (1.37)
Math grade		9.96 (2.01)	10.52 (1.91)	10.79 (1.98)	11.95 (2.78)	11.49 (2.67)	10.06 (3.42)
Cohort 2	330						
Math SCA		5.49 (0.67)	5.45 (0.69)	5.32 (0.73)	5.00 (0.82)	4.76 (0.89)	4.73 (0.94)
Math grade		10.65 (1.36)	10.99 (1.47)	11.87 (1.13)	11.97 (2.00)	11.00 (2.36)	11.13 (2.23)
Cohort 3	423						
Math SCA		5.35 (0.76)	5.35 (0.78)	5.25 (0.82)	4.82 (0.93)	4.66 (1.01)	4.54 (1.07)
Math grade		11.11 (1.54)	12.04 (1.67)	11.99 (1.28)	10.43 (2.26)	9.58 (2.67)	9.92 (2.53)
<b>CDP</b>							
Cohort 1	308						
Aggression efficacy		0.48 (0.32)	0.43 (0.30)	0.42 (0.33)	0.40 (0.33)	0.39 (0.32)	0.32 (0.35)
CBCL		2.64 (0.86)	2.45 (0.90)	2.38 (0.97)	2.14 (1.07)	2.08 (1.32)	1.93 (1.11)
Cohort 2	277						
Aggression efficacy		0.46 (0.30)	0.41 (0.30)	0.38 (0.30)	0.37 (0.32)	0.35 (0.32)	0.33 (0.37)
CBCL		2.45 (0.85)	2.31 (0.95)	2.13 (0.98)	2.07 (1.33)	2.19 (1.35)	2.05 (1.21)

Note. Standard deviations are in parentheses. CAB = Childhood and Beyond Study; CDP = Child Development Project; CBCL = Child Behavior Checklist; SCA = self-concept of ability.

Table 2

Factor Loadings for Math Self-Concept of Ability (CAB), Math Grade (CAB), Efficacy of Aggression (CDP), and Aggression (CDP)

	Point of assessment					
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Math self-concept of ability (CAB)						
How good at math are you?	.776 (.41)	.847 (.29)	.858 (.27)	.903 (.21)	.921 (.18)	.910 (.19)
If you had to list all the students from best to worst in math where are you?	.720 (.49)	.782 (.40)	.829 (.32)	.749 (.48)	.779 (.44)	.779 (.42)
Compared to other subjects how good are you at math?	.490 (.76)	.620 (.62)	.664 (.57)	.839 (.33)	.862 (.30)	.841 (.32)
How well do you expect to do in math this year?	.670 (.56)	.593 (.66)	.710 (.51)	.762 (.46)	.712 (.54)	.829 (.33)
How good would you be at learning something new in math?	.489 (.77)	.430 (.82)	.574 (.68)	.726 (.51)	.796 (.41)	.820 (.35)
Math grade (CAB)	.926 (.14)	.924 (.15)	.942 (.11)	.977 (.05)	.981 (.04)	.982 (.04)
Efficacy for aggression (CDP)						
Pushing kids out of line is hard for you?	.462 (.79)	.514 (.74)	.552 (.70)	.664 (.56)		
Telling a kid to stop changing the channel or you will hit him/her is hard for you?	.723 (.48)	.666 (.56)	.603 (.64)	.750 (.44)		
Telling a kid that he/she had better let you play is hard for you?	.622 (.61)	.661 (.56)	.742 (.45)	.766 (.41)		
Telling a kid that you will hit him/her if he/she does not let you see his/her photos is hard for you?	.608 (.63)	.682 (.53)	.697 (.51)	.708 (.50)		
How easy would it be for you to act like this? (Vignette 1)					.384 (.85)	.653 (.57)
How easy would it be for you to act like this? (Vignette 2)					.602 (.63)	.744 (.44)
How easy would it be for you to act like this? (Vignette 3)					.550 (.70)	.721 (.48)
How easy would it be for you to act like this? (Vignette 4)					.764 (.42)	.707 (.50)
How easy would it be for you to act like this? (Vignette 5)					.766 (.41)	.595 (.65)
How easy would it be for you to act like this? (Vignette 6)					.674 (.54)	.747 (.44)
Child Behavior Checklist (CDP)	.933 (.13)	.933 (.13)	.932 (.13)	.944 (.11)	.945 (.11)	.946 (.11)

Note. Residual variances in parentheses. CAB = Childhood and Beyond Study; CDP = Child Development Project.

Model 6,  $\chi^2_{\text{diff}}(26) = 35.18, p = .11$ , and the  $\chi^2_{\text{diff}}$  test did not significantly differ from the unconstrained model. Thus, we can conclude that the associations between math self-efficacy and math grades did not differ across sex.

Given these findings, we proceeded to estimate the simplex model of relations between math self-efficacy and math grades across grades separately for each cohort with the parameters constrained to be equal across males and females. The resulting parameter estimates are shown in Figure 1 (note that the standardized estimates or correlations are presented opposed to the unstandardized estimates or covariances). The estimates for the synchronous correlations between math self-concept and math grades are of key interest. In order to test if the synchronous correlations vary over time, we conducted a series of model comparisons. First, we compared a model where all synchronous covariances were free to vary across time (Model 7, Table 3) to a model where all synchronous covariances were constrained to be equal across time (Model 8, Table 3). Model 7 (where synchronous covariances were allowed to vary across time) fit the data better than did Model 8,  $\chi^2_{\text{diff}}(15) =$

272.20,  $p < .001$ . Thus, the association between math self-concept and math grades does vary across time. Next, guided by our expectation that the association between beliefs and behaviors would be markedly higher during adolescence than during childhood, we compared a model where all synchronous covariances between Grade 1 and Grade 5 (i.e., childhood) were constrained to be equal and all synchronous covariances between Grades 6 and 12 (i.e., adolescence) were constrained to be equal (Model 9, Table 3) to the more parsimonious Model 8 (Table 3). Model 9 fit the data better than did Model 8,  $\chi^2_{\text{diff}}(3) = 111.66, p < .001$ . As expected, the estimate for the synchronous association during childhood (unstandardized = .51, standardized  $\sim$  .26) was lower than the estimate for the synchronous association during adolescence (unstandardized = 1.94, standardized  $\sim$  .51). Finally, guided by our expectation that the association between self-efficacy beliefs and behaviors would also be markedly higher during late childhood (Grades 3–5) than during early childhood (Grades 1–2), we compared a model where all synchronous covariances between Grade 1 and Grade 2 (i.e., early childhood) were constrained to be equal,



Table 3  
*Childhood and Beyond Model Comparisons*

Model	$\chi^2$	df	Chi-square, <i>p</i> value	CFI	RMSEA	$\chi^2_{diff}$	df diff	<i>p</i> value
Cohort comparison								
(1) Cohort—unconstrained model	3,339.58	1425	< .001	.919	.061			
(2) Cohort—synchronous covariances constrained	3,399.95	1437	< .001	.917	.062	60.37	12	< .001
(3) Cohort—“structural” covariances constrained	3,536.77	1477	< .001	.912	.063	197.19	52	< .001
Gender comparison								
(4) Gender—unconstrained model	2,472.67	942	< .001	.932	.055			
(5) Gender—synchronous covariances constrained	2,480.63	948	< .001	.932	.055	7.96	6	.24
(6) Gender—“structural” covariances constrained	2,507.85	968	< .001	.932	.055	35.18	26	.11
Structural comparison								
(7) Unconstrained	3,339.58	1425	< .001	.919	.061			
(8) All synchronous covariances constrained	3,611.78	1440	< .001	.908	.065	272.20	15	< .001
(9) Constraining 1st–5th versus 6th–12th	3,500.13	1437	< .001	.912	.063	111.66	3	< .000 <sup>a</sup>
(10) Constraining 1st–2nd versus 3rd–5th versus 6th–12th	3,497.86	1435	< .001	.912	.063	2.269	2	.32 <sup>b</sup>

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation.

<sup>a</sup>Model comparison is based on comparing Model 9–Model 8 (i.e., Model 9 fits better).

<sup>b</sup>Model comparison is based on comparing Model 10–Model 9 (i.e., Model 9 and 10 fit equally well).

all synchronous covariances between Grades 3 and 5 were constrained to be equal (i.e., late childhood), and all synchronous covariances between Grades 6 and 12 (i.e., adolescence) were constrained to be equal (Model 10, Table 3) to the more parsimonious Model 9 (Table 3). Model 10 did not fit the data any better than did Model 9,  $\chi^2_{diff}(2) = 2.27$ ,  $p = .32$ . Thus, the synchronous correlations between math self-efficacy and math grades do not appear to differ between early childhood and late childhood.

#### *Changes in Synchronous Correlations Between Efficacy for Aggression and Aggressive Behavior*

Again, before examining whether or not the association between efficacy for aggression and aggressive behavior changed over time, we tested for cohort and sex differences in the measurement model presented in Figure 2. First, we tested for cohort differences. Specifically, a fully unconstrained model (Model 1, Table 4) where all covariances were free to vary across the two cohorts was compared to both a model where just the synchronous covariances were constrained to be equal across the two cohorts (Model 2, Table 4) and a model where all covariances were constrained to be equal across the two CDP cohorts (Model 3, Table 4). The fit of Model 1 (the fully unconstrained model) was not significantly better than either Model 2,  $\chi^2_{diff}(6) = 8.53$ ,  $p = .20$ , or Model 3,  $\chi^2_{diff}(26) = 35.93$ ,  $p = .09$ . Thus, we concluded that the associations between self-efficacy for aggression and aggressive behavior did not differ across the two CDP cohorts. Next, we tested for sex differences by

comparing a fully unconstrained model (Model 4, Table 4) where all covariances were free to vary across sex to both a model where just the synchronous covariances were constrained to be equal across sex (Model 5, Table 4) and a model where all covariances were constrained to be equal across sex (Model 6, Table 4). The fit of Model 4 (the fully unconstrained model) did not fit the data any better than did Model 5,  $\chi^2_{diff}(6) = 8.35$ ,  $p = .21$ , or Model 6,  $\chi^2_{diff}(26) = 28.95$ ,  $p = .31$ . Thus, we concluded that the associations between self-efficacy of aggression and aggressive behaviors (including synchronous correlations) did not differ across sex.

Given these findings, we proceeded to estimate the simplex model of relations between self-efficacy of aggression and aggressive behaviors with the parameters constrained to be equal across sex and cohort. The resulting parameter estimates are shown in Figure 2 (note that the standardized estimates or correlations are presented opposed to the unstandardized estimates or covariances). The estimates for the synchronous correlations between self-efficacy of aggression and aggressive behaviors are of key interest. In order to test if the synchronous correlations vary over time, we conducted a series of model comparisons. First, we compared a model where all synchronous covariances were free to vary across time (Model 7, Table 4) to a model where all synchronous covariances were constrained to be equal across time (Model 8, Table 4). Model 7 (where synchronous covariances were allowed to vary across time) fit the data better than did Model 8,  $\chi^2_{diff}(5) = 49.65$ ,  $p < .001$ . Thus, the association between self-efficacy of

Table 4  
*Child Development Project Model Comparisons*

Model	$\chi^2$	df	Chi-square, p value	CFI	RMSEA	$\chi^2_{diff}$	df diff	p value
Cohort comparison								
(1) Cohort—unconstrained model	1,646.59	896	< .001	.899	.054			
(2) Cohort—synchronous covariances constrained	1,655.18	902	< .001	.899	.053	8.525	6	ns
(3) Cohort—“structural” covariances constrained	1,682.52	922	< .001	.898	.053	35.928	26	ns
Gender comparison								
(4) Gender—unconstrained model	1,669.37	896	< .001	.898	.054			
(5) Gender—synchronous covariances constrained	1,677.71	902	< .001	.898	.054	8.348	6	ns
(6) Gender—“structural” covariances constrained	1,698.31	922	< .000	.897	.054	28.945	26	ns
Structural comparison								
(7) Unconstrained	1,103.86	437	< .001	.909	.051			
(8) Constraining all synchronous covariances	1,153.52	442	< .001	.903	.053	49.653	5	< .001
(9) Constraining K to 3 versus 8–11	1,118.52	441	< .001	.908	.052	34.997	1	< .001 <sup>a</sup>
(10) Constraining K to 2 versus 3 versus 8–11	1,107.35	440	< .001	.909	.051	11.167	1	< .001 <sup>b</sup>

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation.

<sup>a</sup>Model comparison is based on comparing Model 9–Model 8 (i.e., Model 9 fits better).

<sup>b</sup>Model comparison is based on comparing Model 10–Model 9 (i.e., Model 10 fits better).

aggression and aggressive behaviors does vary across time. Next, guided by our expectation that the association between beliefs and behaviors would be markedly higher during adolescence than during childhood, we compared a model where all synchronous covariances between kindergarten and Grade 3 (e.g., childhood) were constrained to be equal and all synchronous covariances between Grades 8 and 11 (i.e., adolescence) were constrained to be equal (Model 9, Table 4) to the more parsimonious Model 7 (Table 4). Model 9 fit the data better than did Model 7,  $\chi^2_{diff}(1) = 34.99$ ,  $p < .001$ , demonstrating that the relations between self-efficacy of aggression and aggressive behavior were similar within childhood and within adolescence but differed between childhood and adolescence. As expected, the estimate for the synchronous association during childhood (unstandardized = .00, standardized  $\sim .01$ ) was lower than the estimate for the synchronous association during adolescence (unstandardized = .06, standardized  $\sim .30$ ). Finally, guided by our expectation that the association between beliefs and behaviors would be markedly higher during late childhood (Grades 3) than during early childhood (kindergarten–Grade2), we compared a model where all synchronous covariances during early childhood were constrained to be equal, all synchronous covariances during late childhood were constrained to be equal, and all synchronous covariances during adolescence where constrained to be equal (Model 10, Table 4) to the more parsimonious model 9 (Table 4). Model 10 fit the data better than did Model 9,  $\chi^2_{diff}(1) = 11.17$ ,  $p <$

.001. As expected, the synchronous association during late childhood (unstandardized = .04, standardized  $\sim .15$ ) was higher than the synchronous association during early childhood (unstandardized = .00, standardized =  $-.02$ ) but lower than the synchronous association during adolescence (standardized = .06, unstandardized  $\sim .30$ ).

## Discussion

Decades of theoretical and empirical developmental research have suggested that a cognitive-developmental shift in middle childhood may account for well-documented differences between cognitive functioning in early and later childhood (Case, 1991; Harter, 1996; Sameroff & Haith, 1986). Among these developmental progressions are several that involve how children think about and judge the behaviors of self and others (Flavell & Miller, 1998; Wellman, Cross, & Watson, 2001). This article presents evidence of a similar developmental progression in links between children’s self-efficacy beliefs and their own behaviors. Although many theoretical models in the social sciences imply that this link should be stable across development (Crick & Dodge, 1994; Fazio, 1995), our results suggest that, at least in the case of some self-efficacy beliefs, this assumption is not tenable.

Underlying this conclusion are our findings that the relation between beliefs in two different psychological domains (self-concept of ability for math, self-efficacy for aggression) and behaviors (grades,

mother reports of child aggression) increases with development independently of the ability to measure the constructs. It is notable that these relations do not differ by sex for either math achievement or aggression where we might have expected differences to exist.

Another strong finding is that these changes replicate in four independent samples. In the case of math achievement, three independent cohorts assessed with the same self-efficacy and behavior measures across time showed the same shift in relations between childhood and adolescence and also displayed very similar relations at the same point in development (grade). The same findings replicated for self-efficacy of aggression and aggressive behaviors with two independent cohorts. Interestingly, when examining aggression beliefs and behaviors, there appears to be an additional significant shift in the relations between early childhood and middle childhood. Thus, there does appear to be an important adolescent change in the relation between self-efficacy beliefs and behaviors for both achievement and aggression beliefs but an additional shift for aggression beliefs in middle childhood. Examining whether this shift is consistent with cognitive-developmental changes that are being seen in early adolescence (Keating, 2004) will be an important avenue of research to pursue in the future.

One of the many challenges in explaining this change is accounting for the possibility that the reliability of the relevant measures may increase across time, thus producing the more stable age-related relations. Previous research has indicated that reliabilities and stability in measurement do increase by age (Davis-Kean & Sandler, 2001; Guay, Marsh, & Boivin, 2003; Trzesniewski et al., 2003), though even young children appear to respond reliably in these domains and can even make complicated differentiations across multiple aspects of the self (Marsh, Debus, & Bornholt, 2005). To rule out measurement unreliability as an alternative explanation for the present findings, our analyses corrected for attenuation by using latent variables with multiple indicators in structural equation models. Furthermore, as can be seen in the Method section, the internal reliability coefficients at both the earliest and the later ages showed adequate to good internal reliability. Similarly, the within-construct correlations across time (i.e., the extent to which beliefs and behaviors correlated with themselves across time) were both sizable and stable, revealing that belief and behavior structures remain relatively stable across time. Thus, the measures across the studies appear to be generally good indicators of the constructs measured, and the

observed changes in the concurrent relations across time are unlikely to be due to changes in the reliability of the measures.

Developmental psychologists have demonstrated that children are capable of certain thoughts and actions well before the classic Piagetian stages would have predicted them (Gelman & Baillargeon, 1983). However, isolating a reliable developmental progression in belief–behavior relations implies both groups have been correct in concluding that the early development of beliefs is a precursor to later knowledge about the self. This knowledge, however, may only become accessible and predictive at a certain point in development (Harter, 1999).

The present findings thus underscore the importance of several methodological issues for future studies of belief–behavior relations. Our data rely on self-report methods of self-efficacy beliefs, and the observed changes in synchronous relations can only be generalized to other self-report findings. Similarly, beliefs other than self-efficacy beliefs should be examined to determine whether belief–behavior relations vary across types of self-beliefs. As Wellman et al. (2001) have found, false beliefs are well established by the age of 5, and thus we would not necessarily expect to see these types of changes in relations across time in tasks related to false beliefs.

Even with these limitations, the findings underscore the importance of several methodological issues for future studies of belief–behavior relations. Questions of how and when in development these phenomena should be measured also remain important for future research endeavors. Should researchers collect self-report data on self-conceptions of beliefs if the purpose is to predict behavior, especially in cross-sectional data collection? The fact that the beliefs data in the present study are self-reports from children, whereas the behavior indicators are school records and reports from parents constitutes both a strength and weakness for interpreting the analyses. Using reports from others or school records avoids the method variance problems that occur when measures of both the independent and the dependent variables come from self-reports of study participants. Experimental studies would potentially permit testing these hypotheses directly and give us additional information on the potential mechanisms underlying the change in the relations between beliefs and behaviors across time.

Despite these remaining challenges, the present evidence of changes across time in the connection between self-efficacy beliefs and behaviors is a potentially significant advance in understanding the nature of belief–behavior relations. That the finding is

robust across multiple belief-behavior domains (achievement and aggression) and using different measures of behavior (grades, parent reports of aggression) strengthens the contribution of the work as a useful beginning point for understanding this phenomenon and its implications for future research on theory and methods. Finally, confirming relatively late increases in stability implies the possibility of intervening with young children at risk for potentially dysfunctional behavior-relevant beliefs. If children's beliefs regarding their academic performance or the appropriateness and effectiveness of certain behavioral strategies are in formation well into the school years, clinicians have a window of opportunity during which interventions directed toward these beliefs might lead to more positive and successful behaviors in the future.

### References

- Achenbach, T. M. (1991). *Child Behavior Checklist and Child Behavior Profile: Cross-Informant version*. Burlington: University of Vermont.
- Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72*, 1409–1425.
- Bandura, A. (1986). *Social Foundation by thought and action*. Englewood Cliffs, NJ: Prentice Hall.
- Case, R. (1991). Stages in the development of the young child's first sense of self. *Developmental Review, 11*, 210–230.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.
- Davis-Kean, P. E., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for young children: A framework for future measures. *Child Development, 72*, 887–906.
- Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678–1683.
- Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Ontogeny of children's self-perceptions and subjective task values across activity domains during the early elementary school years. *Child Development, 64*, 830–847.
- Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Erlbaum.
- Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon (Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology* (5th ed., pp. 851–898). New York: Wiley.
- Gelman, R., & Baillargeon, R. (1983). A review of Piagetian concepts. In J. H. Flavell, M. E. Markman, & P. H. Mussen (Eds.), *Handbook of child psychology, Vol. 3. Cognitive development*. (pp. 167–230). New York: Wiley.
- Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (NIDA Research Monograph No. 142, pp. 13–63). Washington, DC: National Institute on Drug Abuse.
- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*, 124–136.
- Harter, S. (1996). Developmental changes in self-understanding across the 5 to 7 year shift. In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift*. (pp. 207–236). Chicago: University of Chicago Press.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford Press.
- Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior, 14*, 13–24.
- Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72*, 408–419.
- Joreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121–145.
- Keating, D. P. (2004). Cognitive and brain development. In R. J. Lerner & L. D. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 45–84). New York: Wiley and Sons.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Skokie, IL: Rand McNally.
- Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology, 34*, 709–718.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement, 30*, 157–183.
- Marsh, H. W., Debus, R. L., & Bornholt, L. J. (2005). Validating young children's self-concept responses: Methodological ways and means to understand their responses. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 138–160). Oxford, UK: Blackwell.
- Muthen, L. (2005). *Multiple imputation discussion*. Retrieved, October 15, 2007, from <http://www.statmodel.com/discussion/messages/22/381.html>
- Muthen, L., & Muthen, B. (1998–2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthen & Muthen.

- Parsons, J. E., & Ruble, D. N. (1977). The development of achievement-related expectancies. *Child Development, 48*, 1075–1079.
- Piaget, J. (2001). *The psychology of intelligence*. New York: Rutledge. (Original work published 1963)
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley.
- Ruble, D. N., Parsons, J. E., & Ross, J. (1976). Self-evaluative responses of children in an achievement setting. *Child Development, 47*, 990–997.
- Sameroff, A. J., & Haith, M. M. (1986). *The five to seven year shift*. Chicago: University of Chicago Press.
- SAS Institute Inc. (1999). *SAS procedures guide, Version 8*. Cary, NC: Author.
- Schafer, J. L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model. Software for Windows 95/98/NT, Version 2*.
- Schafer, J., & Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545–571.
- Siegler, R. S., & Alibali, M. W. (2004). *Children's thinking* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Spencer, F. H., & Bornholt, L. (2003). A model of children's cognitive functioning and cognitive self concepts. *Australian Journal of Learning Disabilities, 8*, 4–8.
- Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology, 84*, 205–220.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.
- White, S. (1970). Some general outlines of the matrix of developmental changes between 5 and 7 years. *Bulletin of the Orton Society, 20*, 41–57.
- White, S. H. (1996). The child's entry into the "age of reason." In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift* (pp. 17–30). Chicago: University of Chicago Press.
- Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265–310.