



Effects of Developmentally Appropriate Practices on Social Skills and Problem Behaviors in 1st Through 3rd Grades

M. Lee Van Horn , Emilie Karlin & Sharon Ramey

To cite this article: M. Lee Van Horn , Emilie Karlin & Sharon Ramey (2012) Effects of Developmentally Appropriate Practices on Social Skills and Problem Behaviors in 1st Through 3rd Grades, Journal of Research in Childhood Education, 26:1, 18-39, DOI: [10.1080/02568543.2012.633843](https://doi.org/10.1080/02568543.2012.633843)

To link to this article: <https://doi.org/10.1080/02568543.2012.633843>



Published online: 27 Dec 2011.



Submit your article to this journal [↗](#)



Article views: 773



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Effects of Developmentally Appropriate Practices on Social Skills and Problem Behaviors in 1st Through 3rd Grades

M. Lee Van Horn

University of South Carolina, Columbia, South Carolina

Emilie Karlin

American Indian Prevention Coalition, Phoenix, Arizona

Sharon Ramey

Virginia Tech Carilion School of Medicine and Research Institute, Roanoke, Virginia

The guidelines published by the National Association for the Education of Young Children on the use of developmentally appropriate practices (DAP) have, over the last two decades, had an important influence on young children's educational experiences. The efficacy of these guidelines for changing children's outcomes has been examined by only a handful of studies and with mixed results. This study looks at the effects of classroom and school-level use of classroom elements of DAP in 1st, 2nd, and 3rd grades on parent ratings of children's social skills at the end of the year, controlling for ratings of the same construct at the end of the previous year, gender, and ethnicity with a sample of between 1,145 and 2,111 students each year. Differential effects of DAP for males and females and for children of different ethnic backgrounds also were examined. Despite high power, no consistent effects of DAP were observed and no interactions found, suggesting that DAP does not affect parents' ratings of social skills in 1st through 3rd grades. The importance of these results, in light of other work on the effects of DAP, is discussed.

Keywords: developmentally appropriate practices, social skills, problem behaviors, student outcomes, differential effects, effects of teaching

Early childhood education programs tend to fall along a continuum of teaching styles, with those emphasizing didactic teaching of academic skills on one end, and those offering a more "child-centered" approach on the other. The child-centered approach has received considerable attention with the development of the popular developmentally appropriate practices (DAP) guidelines

Submitted September 13, 2009; accepted April 6, 2011.

Preparation of this article was supported by research grant R01 HD043400-01A1, awarded to the first author by the National Institute of Child Health and Human Development.

Address correspondence to M. Lee Van Horn, Department of Psychology, University of South Carolina, Columbia, SC 29208. E-mail: vanhorn@sc.edu

by the National Association for the Education of Young Children (NAEYC; Bredekamp, 1987; Bredekamp & Copple, 1997; Copple & Bredekamp, 2009). The DAP guidelines portray the child as an active participant in the learning process, emphasizing play as a means of learning, and the importance of using teaching methods appropriate to the developmental stage of the class as a whole and to the individual needs of each child (Copple & Bredekamp, 2009). The DAP guidelines are widely endorsed by early childhood educators partly because they make intuitive sense and offer concrete examples of appropriate classroom practices.

Although many academics and educators have embraced the DAP guidelines, others have been more critical. There are three major criticisms of DAP: (1) its guidelines purport to provide a set of global best practices yet disregard diversity and cultural differences, (2) limited support exists for its theoretical base, and (3) there is little empirical evidence of its efficacy in improving child outcomes. Although the revised version of the DAP guidelines addresses cultural issues more thoroughly than the initial version, critics argue that the general principles of DAP reflect middle-class White values rather than those of more collectivistic cultures (Hsue & Aldridge, 1995; Smith, 1996). Academics also have debated the value of having global written guidelines for best practices (Charlesworth, 1998a, 1998b; Lubeck, 1998a, 1998b). Furthermore, there are very few empirical studies that specifically evaluate the effects of DAP on children from different cultures (Van Horn, Karlin, Ramey, Aldridge, & Snyder, 2005).

The NAEYC guidelines cite Piaget and Vygotsky as theoretical influences on DAP, but critics argue that DAP guidelines focus mostly on the stage-like system of development proposed by Piaget and place significantly less emphasis on the social and cultural influences on development suggested by Vygotsky (Smith, 1996; Walsh, 1991). The DAP guidelines state that it is crucial to incorporate information about children based on their developmental level and on their individual characteristics, but much more time is spent discussing appropriate developmental levels (Aldridge, 1992). This may be problematic, because research has shown limited support for the notion of stage-like development (Gelman & Baillargeon, 1983). Indeed, critics of DAP stress the importance of Vygotsky's zone of proximal development, in which children are challenged to learn by engaging in tasks beyond their current capabilities (Smith, 1996; Walsh, 1991).

Proponents of DAP indicate that DAP is expected to create a rich classroom environment that will support cognitive and social development (Elkind, 1986, 1987). The DAP guidelines list both academic skills, such as math and reading, and psychosocial skills, such as getting along with other children and other related classroom behaviors, as important aspects of developmentally appropriate early childhood education programs (Copple & Bredekamp, 2009). Despite their widespread acceptance and use, however, surprisingly little empirical research has considered the impact of DAP on either academic or psychosocial outcomes (Van Horn et al., 2005).

EFFECTS OF DAP ON ACADEMIC OUTCOMES

In a recent review of studies on the effectiveness of DAP, Van Horn and colleagues (2005) found no consistent impact of DAP on academic outcomes. Several researchers have reported positive results of DAP that are consistent across a variety of academic subjects, such as math, reading, and science (Huffman & Speer, 2000; Marcon, 1992, 1999). Other studies have reported more mixed results, in which DAP had a positive impact for some subjects but not others (Burts et al., 1993; Marcon, 1993). Surprisingly, numerous studies have found that DAP has a negative impact on academic outcomes, particularly reading (Hyson, Hirsh-Pasek, & Rescorla, 1990;

Jones & Gullo, 1999; Stipek, Feiler, Daniels, & Milburn, 1995; Stipek et al., 1998). Of particular interest, the study with the largest sample size and greatest power failed to find a consistent impact of DAP on academic outcome variable for students in 1st, 2nd, or 3rd grade (Van Horn & Ramey, 2003).

EFFECTS OF DAP ON PSYCHOSOCIAL OUTCOMES

Although there is limited evidence that DAP has a positive impact on academic outcomes, positive effects on social skills may justify their continued use. Social skills may include areas of competence, such as cooperation and peer relationships, and problem behaviors, such as disruptiveness, withdrawal, or hyperactivity (Gresham & Elliott, 1988, 1990). Children who engage in more prosocial behaviors, such as rule following and friendliness, tend to be rated more positively by their peers; children who engage in more negative behaviors, such as aggression and disruptiveness, tend to be rated more negatively by their peers (Coie, Dodge, & Kupersmidt, 1990). Children who are rated negatively by their peers are more likely to experience a variety of later problems—both school-related problems, such as truancy or early dropout (Woodward & Fergusson, 2000)—and social problems, such as delinquency (Roff, Sells, & Golden, 1972). These children are also likely to continue to have difficulties with peer relationships as they get older (Ladd & Troop-Gordon, 2003). Improving children's social skills at a young age may, therefore, have important implications for their later development.

The current study assesses the effects of DAP on social skills. The available research on the effects of DAP on social outcomes is more consistent and positive than the research on academic outcomes (Van Horn et al., 2005). Much of the research on psychosocial outcomes has focused on stress and anxiety. Studies have consistently demonstrated that children in DAP classrooms experience significantly less stress and anxiety about school than children who are not in DAP classrooms (Burts et al., 1992; Burts, Hart, Charlesworth, & Kirk, 1990; Hart et al., 1998; Stipek et al., 1995). These results are consistent for kindergarten and preschool students. Children in DAP classrooms have also scored significantly higher on other measures of positive psychosocial attributes, such as high expectations for success, greater pride in their accomplishments, greater creativity, and more positive attitudes about school (Burts et al., 1992; Burts et al., 1990; Hart et al., 1998; Hirsh-Pasek, Hyson, & Rescorla, 1990; Stipek et al., 1995; Stipek et al., 1998). Children in classrooms that primarily use didactic teaching strategies demonstrate greater levels of negative affect, more noncompliance to teacher requests, and greater dependency on others than students in DAP classrooms (Stipek et al., 1995; Stipek et al., 1998). Only a few studies have failed to find positive impacts of DAP on psychosocial outcomes. Specifically, one study of self-competence found that only one of four scales, peer acceptance, was related to levels of DAP in the classroom (Jambunathan, Burts, & Pierce, 1999). Two studies of adaptive skills, including such domains as communication, socialization, and activities of daily living, found no differences between children in DAP classrooms and children in classrooms that were not DAP (Marcon, 1992, 1993). Taken together, the results from numerous studies suggest that DAP helps children develop a variety of psychosocial skills.

Although the findings for psychosocial outcomes are more consistent, results of many existing studies of DAP may be compromised by methodological flaws. Perhaps the most serious of these flaws is the widespread failure to use analyses appropriate to nested designs—that is, using a

design that examines multiple children within a single classroom without the use of appropriate analytical methods (Van Horn et al., 2005; Van Horn & Ramey, 2004). Commonly used analyses, such as regression and ANOVA, assume that each observation in a data set is independent of the others, conditional on the variables in the model. In a nested design, children are linked by classroom, and typically the outcomes for children within a classroom are more similar to each other than those of children from other classrooms or schools (Tabachnick & Fidell, 1996). When analyses fail to account for nesting, standard errors will be underestimated and hypothesis tests will be overly liberal (Bryk & Raudenbush, 1992; Heck & Thomas, 2000; Hox, 1995, 1998).

The extent to which the violation of this assumption affects results depends on the degree to which outcomes are more similar within classrooms than between classrooms (measured by the intraclass correlation coefficient [ICC]) and the number of children assessed per classroom (Murray, 1998). Many of the studies of DAP include only a small number of classrooms, with a large number of students in each classroom; and child outcomes are often fairly highly clustered within classrooms and schools (Raudenbush & Bryk, 2002; Van Horn & Ramey, 2003). Thus, analyses often result in serious violation of the assumption of independence, potentially invalidating the results. In a recent review of the literature on the effects of DAP, Van Horn and colleagues (2005) used Monte Carlo simulations to evaluate the amount of bias present in previous research on the effects of DAP. They found that only five of 23 results reported in the literature that found significant effects of DAP would have been likely to be significant under realistic levels of ICCs if the appropriate analyses had been conducted. Only three of the five effects that remained significant when appropriate analyses were conducted found positive effects of DAP on social skills. The current study will further investigate the effects of DAP on children's social skills.

DIFFERENTIAL EFFECTS OF DAP

An additional area of research on the effects of DAP involves investigating the possibility that effects may vary across individual children. Existing research on differential effects of DAP across ethnicity and gender is sparse and has found equivocal results. Some evidence suggests that boys in classrooms lower in DAP experience more stress than boys in high-DAP classrooms, whereas no such differences exist for girls (Burts et al., 1992; Burts et al., 1990). Another study found that boys in classes that emphasized socioemotional development performed better on a measure of adaptive skills than boys in classes that emphasized academic development (Marcon, 1993). For girls, however, this pattern was reversed, with girls in classrooms that emphasized academic development performing better than girls in classes that emphasized socioemotional development (Marcon, 1993). Still other researchers, however, have found no gender differences (Marcon, 1999; Stipek et al., 1995). There is even less information about the differential effects of DAP across ethnicities. One study reported no differences between African American and White children in DAP classrooms on measures of stress, but found that African American and White children in less-DAP classrooms reacted stressfully to different activities, and that they tended to be involved in different activities as well (Burts et al., 1992). It is important to learn if different groups of children are affected differentially by exposure to DAP; if so, then some groups may be placed at a disadvantage by its use.

Although the DAP guidelines propose implementing DAP in classrooms serving children from birth through age 8 (Bredekamp & Copple, 1997), only two of the studies mentioned above examined effects of DAP in 1st grade. One study found that children exhibited greater stress when engaging in computer activities that were less DAP than activities that were more DAP (Ruckman, Burts, & Pierce, 1999). Another study found that children in DAP classrooms demonstrated higher social skills than children in less-DAP classrooms (Jones & Gullo, 1999). No previous research has evaluated the effects of DAP on psychosocial outcomes in 2nd or 3rd grades, although one study looking at the measurement of DAP did find that though the range of DAP remained high across grades, there was a significant but small decrease in average levels of DAP from 1st through 3rd grades (Buchanan, Burts, Bidner, White, & Charlesworth, 1998). A review of the literature suggests that methodologically sophisticated research evaluating the effects of DAP in general and the effects of DAP on different groups of students is needed, and that evaluations of the effects of DAP on students at the upper end of the age range for which the guidelines are intended are especially important (Van Horn et al., 2005).

AIMS

This study had two major aims: (1) to examine the extent to which DAP implemented in 1st-through 3rd-grade classrooms is associated with parent ratings of children's social skills and problem behaviors and (2) to examine the extent to which the effects of classroom DAP on social skills and problem behaviors vary across child ethnicity and gender. Analyses use multilevel and longitudinal methods, which consider the nested data structure and eliminate bias in estimates of standard errors. Parent (rather than teacher) ratings of social skills are used in these analyses for theoretical and methodological reasons. Theoretically, the use of parent ratings allows for an assessment of whether the effects of DAP generalize beyond the classroom. Further, DAP classrooms are structured differently than non-DAP classrooms, allowing more opportunities for social interaction. It is possible that students' social behaviors in the classroom differ because of structural differences, but that those differences do not reflect differences in underlying levels of social competence. Using parent ratings, which are largely independent of classrooms, we are able to examine whether children in DAP classrooms differ in social skills when in environments that, on average, should not be related to DAP. Second, because the DAP guidelines emphasize positive assessments of students, using teacher reports of social skills introduces a potential bias such that teachers conforming more to the DAP guidelines may give higher ratings because of their use of DAP. In the current study, we used a well-established parent-rated measure of social skills and problem behaviors as the outcome of interest.

METHOD

Data for this study were collected as part of the National Head Start Public School Early Childhood Transition Demonstration Project, a 6-year longitudinal intervention trial that followed the progress of former Head Start children and selected classmates from kindergarten through 3rd grade. Two consecutive cohorts in 30 sites were enrolled. The U.S. Congress legislated and funded this study to test the proposition that providing comprehensive Head Start-like

services to children and families from kindergarten through the 3rd grade would improve former Head Start children's academic, social, and health outcomes (Ramey et al., 2001).

Schools that agreed to participate in the study were randomly assigned to either a transition demonstration group or a comparison group (for results from the overall study, see Ramey et al., 2001). Preliminary analyses for the current study examined the extent to which use of DAP differed between randomly assigned transition demonstration versus comparison schools, as well as other schools to which children moved over the course of their first 4 years in public school. Analyses examined whether classrooms in the transition demonstration schools had higher levels of DAP than those in the comparison schools (random effects models were used, with classrooms nested within schools and treatment modeled as a school level variable) for the 1st, 2nd, and 3rd grades. Results indicate higher DAP scores only in the first year of this study and that the effect size was quite small. This study focused on how naturally occurring variation in DAP relates to changes in student social skills. The entire sample, including students in transition and comparison schools, is included in the analyses.

Participants

Participants included children, their classrooms, and their schools. Outcomes were measured at the child level, and the primary predictors—DAP factor scores—at the classroom and school levels. The schools cover every major geographic area of the United States, include rural and urban areas, and are ethnically diverse as a whole. After the study began, principal investigators from the local sites realized the limitations of the tool being used to measure DAP, the Assessment Profile (Abbott-Shim & Sibley, 1992; Abbott-Shim, Sibley, & Neel, 1998). The Assessment Profile was designed to measure classroom environments in preschool settings rather than to assess DAP in elementary grades. Accordingly, a new rating tool, called A Developmentally Appropriate Practices Template (ADAPT), was developed, for which Margo Gottlieb assumed primary authorship (Gottlieb, 1995, 1997). ADAPT was administered in three consecutive years, yielding data from 401 classrooms within 168 schools for 1st grade, 821 classrooms in 277 schools for 2nd grade, and 922 classrooms in 279 schools for 3rd grade.

Once schools were randomly assigned to a condition, former Head Start children entering kindergarten were recruited to participate in the study. Two cohorts of children were recruited in two consecutive years. Several sites included a sample of non-Head Start children as well. Recruitment of non-Head Start children varied across sites (e.g., some sites recruited children on the same poverty level, whereas other sites recruited a random sample of children), and not all sites followed this sample throughout the remainder of the study. In each grade, between 30% and 31% of the children had not attended Head Start. The measure of DAP used in these analyses was included in 3 years of the study, and data on the outcome measure from the spring of the previous year is used as a baseline measure.

The 1st-, 2nd-, and 3rd-grade analyses include children who had parent ratings of the Social Skills Rating System (SSRS) for that grade and whose classrooms had DAP ratings. Thus, the samples differ for each grade. Further, children identified in school records as having an individualized education plan (IEP), except those identified as having a speech and language impairment exclusively, were excluded from these analyses. This was done because the SSRS has been documented as performing differently for children with disabilities (Gresham & Elliott, 1988, 1990);

thus, this definition of a disability most closely matches the definition provided by the authors of the measure. The final sample included 1,145 children in 1st grade, 2,003 children in 2nd grade, and 2,111 children in 3rd grade. Of those, 36% of the children in the 1st grade were White, 33% were African American, 15% were Hispanic, and most of the remaining specified “other” (usually indicating that the child’s parents identified the child with two or more different ethnic/cultural identities). The sample was 52% male and 48% female. These proportions were similar across all study years.

Measures

DAP was measured using ADAPT (Gottlieb, 1995).¹ ADAPT consists of 19 items (the instrument is copyrighted and cannot be duplicated in full here) derived from an item analysis of the 1987 NAEYC guidelines, plus a “multiculturalism” item that was added to address what was widely perceived to be one of the most prominent drawbacks of the guidelines. The ADAPT items are rated on a 5-point scale, with anchors for each point appropriate to DAP guidelines for each item (Gottlieb, 1997). Table 1 shows examples of items from each ADAPT scale, including

TABLE 1
Example Items for Each ADAPT Scale

<i>Scale/Item Label</i>	<i>Low DAP Anchor (1)</i>	<i>High DAP Anchor (5)</i>
<i>Integrated Curriculum Scale</i>		
Curricular-driven literacy development	Emphasis on basal with phonics and decoding skills, spelling, and mechanics in isolation	Literacy integrated across content areas with literacy materials of social relevance
Multiculturalism	No evidence of linguistic and cultural diversity in classroom or in instructional practices	Multicultural issues integrated into curriculum, instruction, and school, leading to social action
<i>Social/Emotional Emphasis Scale</i>		
Learning is teacher facilitated	Teacher directs all learning and maintains locus of control; formal relationship with children	Teacher facilitates learning, supports children’s decisions, and advocates for each child on his/her behalf
Children are mostly self-regulated	Children lack self-control; unaware of or do not pay heed to rules	Children engage in peer negotiation or conversation to resolve issues, or self-regulation is strongly evident
<i>Child-Centered Approaches Scale</i>		
Children actively interact with materials	Children do not have access to classroom materials outside of textbooks/workbooks	Children encouraged to choose and interact with materials to create and problem-solve
Physical space and arrangement is inviting	Physical arrangement limited to rows of desks, textbooks, commercial displays, and workbooks	Learning centers inviting to children, with variety of real objects; displays of children-generated work

Note. DAP = Developmentally Appropriate Practice; ADAPT = A Developmentally Appropriate Practices Template. ADAPT item anchors are under copyright and reprinted with permission. The full scale has separate anchors for scores 2, 3, and 4, and also has supplemental information for raters for each item. The measure was written by Margo Gottlieb. For further information about this measure, contact Dr. Gottlieb, mgottlieb@cntrmail.org

the lowest and highest anchors. Each criteria for a lower anchor must be met before a higher rating can be given. Although only 19 items are included, each item provides a detailed assessment of that content area. Experts familiar with the NAEYC guidelines reviewed each item for content validity, and modifications were made as needed (Gottlieb, 1997).

In the initial development of the ADAPT, inter-rater agreement was reported to average .69 across items using raters with no previous training in the measure or experience with DAP (Gottlieb, 1997). Before the instrument was used in the national evaluation, a separate test of inter-rater reliability was conducted by two graduate students familiar with the instrument and construct being measured. The students observed and rated the same 30 classrooms. Item-specific kappas ranged from .60 to .80, which is generally in the acceptable range (Suen & Ary, 1989).

For the current study, the authors of the ADAPT tool trained the people overseeing evaluations at each site, who, in turn, trained the raters to a level of 80% agreement or better. Although achievement of this standard could not be verified empirically, reliability can be assessed through a group of teachers who received ADAPT ratings two or more years in a row. Correlations for the same classroom between the first and second years ranged from .52 to .54 for the three ADAPT scales, based on 323 classrooms. For the second and third years, the correlations ranged from .38 to .45 for 476 classrooms; and for the first and third years, correlations ranged from .30 to .50 for 79 classrooms. These correlations represent a lower bound on the reliability of the ADAPT subscale scores. Other factors besides unreliability, such as changes in teachers' use of DAP from one year to the next, would also lower these correlations. Each spring, trained observers spent a total of 2 hours (in 15-minute intervals spread over one day) observing each classroom and scoring the ADAPT and the Assessment Profile, a measure of classroom quality. The use of 15-minute intervals throughout the day helped to ensure that a wide range of activities was observed, which served to increase the degree to which the ratings were representative. All possible response options (1 to 5) were observed across all grades. With one exception, the mean for all items, across all grades, was in the range of 2.11 to 2.93 on a scale of 1 to 5, with standard deviations between .68 and 1.16 and skewness and kurtosis under 1 (details for all items across 1st and 2nd grade can be found in Van Horn & Ramey, 2004). The one exception was the item on multiculturalism, which had a lower mean (1.65) and high skewness (1.58) and kurtosis (2.52). Thus, except for multiculturalism, there was good evidence for a wide range of DAP in these classrooms, with the means being close to the center of the scale. Across classrooms, the entire range of the multiculturalism item was observed as well—there were just far fewer classrooms with high levels of multiculturalism. There was some evidence for small decreases in DAP from 1st to 3rd grades, while overall DAP decreased by .2 standard deviations. This is similar to the results of Buchanan et al. (1998), whose study found mean levels of DAP in about the center of the possible values on a teacher-rated scale, with significant but small decreases in DAP as grade increased.

To validate the ADAPT as a measure of DAP, separate analyses were undertaken, which included a qualitative assessment of the ADAPT to examine the extent to which it assesses the 1997 DAP guidelines, as well as a psychometric analysis to establish the reliability of the measure and the extent to which it assesses different components of DAP (Van Horn & Ramey, 2004). The qualitative analyses found that three of the six components of DAP identified in the guidelines were measured by the ADAPT: (1) creating a caring community of learners, (2) teaching to enhance development, and (3) constructing appropriate curriculum. The three DAP areas not assessed by ADAPT were (1) assessing children's learning and development, (2) establishing

reciprocal relationships with parents, and (3) policies. It is not surprising that the latter three components of DAP are not assessed by the ADAPT, as these areas are not easily assessed by classroom observers and may take place outside of the classroom. Note, also, that because ADAPT measures global aspects of classroom-level DAP rather than specific practices that may be implemented under the guidelines, this has implications for the interpretation of the results. Effects of the ADAPT should be seen as effects of global implementation of DAP in classrooms.

Further psychometric analyses of the reliability and factor structure of the ADAPT were conducted using multilevel structural models. The psychometric analyses revealed three factors at the classroom level—Integrated Curriculum, Social/Emotional Emphasis, and Child-Centered Approaches—that were stable across time and treatment conditions (Van Horn & Ramey, 2004) and comprised 17 of the 19 items assessed by the ADAPT. Integrated Curriculum rates the extent to which the classroom instruction complies with the guidelines, the classroom has a comprehensive purpose, learning is integrated, multiple strategies are used to promote cognitive development, and time is flexible to allow integration. Social/Emotional Emphasis involves the degree to which positive social behavior is enforced, social emotional support is provided, the climate is harmonious, and children work together to regulate their own behavior. The Child-Centered Approaches category addresses the extent to which learning is child-initiated, cooperative, and interdependent, materials are created by children, and children are allowed to form flexible groupings. The ADAPT was also found to be moderately related to the Assessment Profile, another measure of classroom environment, providing evidence for the criterion validity of the instrument.

Social skills were measured using parent rating forms of the SSRS (Gresham & Elliott, 1990). The social skills items were administered to parents in all years of the study, but the problem behavior items were only included in 2nd and 3rd grade. Because of concerns about parent literacy, these data were collected through in-person interviews by a trained interviewer with each parent. Psychometric analyses show that the factor structure of the SSRS is similar to that proposed by the initial authors of the measure (Van Horn, Atkins-Burnett, Ramey, Snyder, & Karlin, 2007). These analyses confirmed the existence of two second-order factors: social skills and problem behaviors. The Social Skills Scale is composed of four subscales, Cooperation, Assertion, Responsibility, and Self-Control. The Problem Behaviors Scale is composed of three subscales: Externalizing Behaviors, Internalizing Behaviors, and Hyperactivity. Further analysis demonstrated that the measurement model for social skills is the same for boys and girls, and for African American, White, and Hispanic children. These analyses also examined longitudinal invariance and found that with some modification of the measure, it could be made invariant across the grade levels represented in this study.

Data Analyses

Missing data. Little data were missing at the classroom level, except for those classes where the ADAPT was not administered, or for child demographic variables. However, nearly one half (48%) of the children were missing parent SSRS ratings for at least one study year, and an additional 28% were missing some items for a given year. Multiple imputation techniques (Little & Rubin, 1987; Schafer, 1997a, 1997b), which make the assumption that data are missing at random (MAR) with respect to the variables in the imputation model, were used to impute missing

parent responses for those children from classrooms with ADAPT ratings. Ten imputations were used and the fraction of missing information obtained from each analysis ranged from .05 to .28 for the effects of DAP.

Multiple imputation involves making multiple estimates of missing data points, based on the values predicted from available data. Each estimate consists of a predicted value derived from a random estimate of the population regression line with the addition of a random error term (Graham, Cumsille, & Elek-Fisk, 2003; Schafer, 1997a). Following the design of this study, imputation was conducted separately for the analyses conducted to assess the effects of DAP in the 1st, 2nd, and 3rd grades, including only students, classrooms, and schools that were present at either the baseline or follow-up time point for that grade. Following imputation, analyses are conducted separately for each imputed data set, and then parameter estimates across the analyses are averaged and standard errors are combined, taking into account the variance in the parameter within each analysis and the variability between imputed data sets. In this study, 10 data sets were imputed using NORM (Schafer, 1997c) and the results combined using SAS PROC MIANALYZE (SAS Institute Inc., 2001a). MIANALYZE uses a technique (Yuan, 2000) for computing approximate error degrees of freedom that limits the maximum degrees of freedom, based on the maximum in a complete data set. Estimates of variance components were obtained by averaging estimates across the analyses; standard errors for these variance components are provided, but great caution is urged when using them, given that the accuracy of these standard errors has been questioned (Singer, 1998).

Measurement. Social skills, problem behaviors, and DAP scores in these analyses were taken from factor scores derived from measurement models previously reported for these constructs. Items were imputed at the item level and then moved into Mplus V3.11 (Muthén & Muthén, 2004). For the social skills measure, factor scores were derived for the baseline and posttest measures for each grade, using a model in which the factor loadings and thresholds were constrained to be equal across years. Polychoric correlations and weighted least squares mean and variance adjusted (WLSMV) estimation were used to account for the ordinal and skewed nature of these data. This method assumes measurement invariance over time, an assumption verified in previous research (Van Horn et al., 2007). For the ADAPT, factor scores were created for each grade following the previously developed measurement model for the ADAPT, which includes three within-classroom factors and one broad school-level factor (Van Horn & Ramey, 2004). The within-classroom factors capture specific aspects of classrooms within schools, whereas the broad school-level factor captures the degree to which DAP is consistently utilized within a school. Again, these analyses utilized polychoric correlations and WLSMV estimation. The factor scores created in Mplus were then standardized at their respective levels to allow easier interpretation of effects.

Model specification. Analyses in this study used a simple repeated-measures design, because the few previous studies that describe changes in social skills found that changes did not follow linear or curvilinear trends. Analyses were conducted separately, examining the effects of DAP in 1st, 2nd, and 3rd grades on social skills. For each model, assessments of social skills at the end of the school year are used as the outcome measure, and assessments at the end of the previous school year are entered as a predictor. Gender and ethnicity (effect coded so that African American, White, and Hispanic students are compared to the grand mean across all groups) are included. Thus, classroom and school DAP is used as a predictor of changes in social skills after

accounting for social skills from the previous year. This study includes two cohorts of children in consecutive school years; thus, one teacher could teach a class from each cohort in consecutive years. To account for this nesting of classes within teachers, we conducted analyses with four levels of nesting including teacher as a level; however, the results showed that for all models, the random effects for teacher were not significant and were typically estimated at 0 or unestimable. Thus, the random effect for teacher was dropped in the final analyses and three levels of nesting are included. At Level 1, children's social skills or problem behaviors scores are predicted by baseline values and by the child-level covariates gender and ethnicity. At Level 2, children are nested within classrooms and classroom-level DAP is entered as a predictor of child outcomes. At Level 3, classrooms are nested within schools and school level DAP is included as a predictor. This model, which uses notation described by Bryk and Raudenbush (1992), is:

Level – 1 : Child

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(Y_{Baseline}) + \pi_{jk}(Covariates) + e_{ijk}$$

Level – 2 : Classroom

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(Instruction) + \beta_{02k}(Prosocial) + \beta_{03k}(Interaction) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{jk} = \beta_{.0k}$$

Level – 3 : School

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(ADAPT) + u_{00k}$$

$$\beta_{01k} = \gamma_{010}$$

$$\beta_{02k} = \gamma_{020}$$

$$\beta_{03k} = \gamma_{030}$$

$$\beta_{10k} = \gamma_{100}$$

$$\beta_{.0k} = \gamma_{.00}$$

$$where : e_{ijk} \approx N(0, \sigma^2), r_{0jk} \approx N(0, \tau_r), u_{00k} \approx N(0, \tau_u)$$

(1)

Y is the outcome variable for child i in classroom j in school k . $Y_{Baseline}$ is the pretest SSRS rating for that child for a given year, and *Instruction*, *Prosocial*, and *Interaction* are the Integrated Curriculum, Social/Emotional Emphasis, and Child Centered Approaches factor scores for the year under study. Thus, π_{0jk} is adjusted social skills or problem behaviors scores for each child after taking into account baseline values and the individual-level covariates. Only the intercepts were modeled as randomly varying across each classroom and school; however, separate models were run, examining the impacts of allowing the pretest

covariates to randomly vary. No differences in the effects of DAP were found for any of these models.

A model-building process was used for these analyses, such that preliminary models were run on one imputed data set to examine model assumptions and correct specification of the random components of the model. As part of this process, the student, classroom, and school-level residuals for the model were plotted against the baseline covariate and the DAP predictors to test the assumptions of homogeneity of variance and that the relationships being modeled were linear. This is an important part of the analyses, because it assesses whether the effects of DAP may occur at only high or low levels of DAP. Lowess lines were fit to these residual plots to allow diagnosis of any nonlinearity in the residuals. For the DAP predictors, no substantial nonlinear trends were observed, indicating that the effects of DAP are consistent across all levels of DAP. One model, in which 2nd-grade problem behaviors was the outcome and 1st-grade social skills were used as the baseline score (because no 1st-grade problem behaviors score was available), did demonstrate a curvilinear relationship with one inflection point. Thus, the square of 1st-grade social skills was also added to the model as a predictor and this eliminated the non-linearity.

All models were analyzed using PROC MIXED (SAS Institute Inc., 2001b), as illustrated by Singer (1998). The parameter estimates, standard errors, covariance matrixes of the parameters, variance components, and *F* tests for omnibus tests when more than one degree of freedom was present (for ethnicity) were output to separate files. The parameter estimates were combined using MIANALYZE, and the variance components and *F* tests were combined using the means procedure. Therefore, all parameter estimates and standard errors reported are accurate, but omnibus *F* tests will only have approximate degrees of freedom for the error term, based on the lowest estimate of the error degrees of freedom for the different effect-coded variables composing the omnibus test.

Analyses were run separately for social skills and (in 2nd and 3rd grade) problem behaviors for each year. Secondary models were then run to examine the interaction of DAP with gender and ethnicity. It would have been preferable to reduce the number of tests conducted by assessing the effects on the five outcomes simultaneously using multivariate techniques. Also, it would be preferable to assess the effects of DAP across all three years simultaneously using a cross-classified design (Goldstein, 1999), rather than separately for each year. However, both of these techniques greatly increase the complexity of an already complex design by adding random coefficients, and they are prohibitive computationally and by sample size constraints, even with large data sets. The two foci of these analyses are the effect size of the predictors and interactions of interest, and the pattern of findings across different outcome variables and years. Because a large number of tests were conducted, interpreting a single finding without other findings supportive of a meaningful pattern would be misleading.

RESULTS

Initial descriptive analyses evaluated the number of students per classroom, per teacher, per school, in each year, and examined the distributions of the DAP, social skills, and problem behaviors factor scores. The distribution of the factor scores is close to normal for all variables.

Effects of DAP in 1st Grade

Initial analyses used a covariate-only model to assess the impact of DAP on social skills (see Table 2). This model shows a strong effect of social skills at the end of kindergarten, on social skills at the end of 1st grade, a fairly weak effect of gender on social skills (females are .11 standard deviations higher than males), and no effects of ethnicity. The next model included the same covariates and added the school-level DAP factor score as a school-level variable. This represents the impact of being in a school in which there are, on average, higher levels of classroom DAP. The effect of school-level DAP on social skills is essentially 0 with a slight increase in the school-level variation, indicating that school-level DAP does not account for any variance in social skills. The third analyses added classroom-level measures of DAP into the second model. Again, the results from Table 2 (note that these models are estimated sequentially and each model includes all the effects from the previous models as well) illustrate that there are essentially no classroom-specific impacts of DAP on social skills in 1st grade. The variance components remain unchanged, indicating that the effect size for these variables is essentially 0. It is important to note that for each of these analyses, plots of residuals with Lowess best fit lines were examined to check for nonlinear relationships among the variables. These plots showed no differential impacts of high versus low DAP. Also, since high multicollinearity between the predictors could potentially hide the relationships of these measures of DAP with outcomes, preliminary models separately assessed effects of each DAP component, finding no significant effects for any of these models.

Final analyses examined whether there were different impacts of DAP on change in social skills in 1st grade for boys and girls or for African Americans, Whites, Hispanics, or other ethnic groups. To do this, analyses examined the interaction of DAP with gender and ethnicity in two separate models. For gender, the overall difference between males and females remained almost unchanged (see Table 2), and there is no evidence of any interaction between gender and DAP in predicting change in outcomes. Next, multiple degree of freedom *F* tests were conducted to assess the differential impact of DAP across ethnic groups. These tests were all found to be nonsignificant (see Table 3), indicating no evidence that the effects of DAP on 1st-grade social skills vary across ethnicity.

Effects of DAP in 2nd Grade

The next set of analyses examined the effects of DAP on social skills and problem behaviors from the end of 1st grade to the end of 2nd grade. Two differences emerged between these analyses and those for 1st grade. First, both cohorts were assessed in 2nd grade and so, in many cases, the same teacher taught the consecutive cohorts. To account for potential dependency in the data, four-level models were estimated in which classrooms were nested within teachers who were, in turn, nested within schools. However, the random intercepts for teachers were, in all cases, estimated to be 0, indicating little variability within teachers across year. Consequently, though the models were first estimated as four-level models, the fourth level was subsequently dropped, and they are reported here as three-level models, as no differences in the effects of DAP were seen between the three- and four-level models. The second difference is that problem behaviors were assessed at the end of 2nd grade but not at the end of 1st grade. Therefore, the

TABLE 2
Effects of DAP on Social Skills in 1st Grade

<i>Model/Effect</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>p</i>
Covariates Only				
Kindergarten social skills	.619	.027	186	.00
Female	.115	.049	419	.02
African American	-.057	.047	257	.23
White	.000	.043	289	1.00
Hispanic	-.058	.066	62	.38
τ_π	.038	.016		
τ_β	.000	.022		
σ^2	.585	.027		
School DAP				
School-level DAP	.001	.032	101	.98
τ_π	.039	.016		
τ_β	.000	.022		
σ^2	.584	.027		
Classroom and School DAP				
School-level DAP	.001	.032	104	.97
Integrated curriculum	-.008	.054	231	.88
Social/emotional emphasis	.034	.065	225	.60
Child-centered approaches	-.023	.069	121	.74
τ_π	.040	.016		
τ_β	.001	.023		
σ^2	.585	.027		
Interaction of DAP With Gender				
Female	.118	.049	402	.02
× School-level DAP	-.003	.051	291	.95
× Integrated curriculum	.074	.098	343	.45
× Social/emotional emphasis	-.002	.128	174	.99
× Child-centered approaches	-.055	.120	383	.65
τ_π	.040	.006		
τ_β	.000	.023		
σ^2	.587	.027		

Note. DAP = Developmentally Appropriate Practice. Degrees of freedom differ because they are estimated *df* from multiple imputation, τ_π is the school-level variance of the intercept, τ_β is the classroom-level variance, and σ^2 is the student-level residual variance.

analyses in which problem behaviors were an outcome includes end-of-1st-grade social skills and its square as predictors to approximate the repeated-measures analyses. Social skills squared are used because examination of model residuals suggested a curvilinear relationship between the two. Thus, social skills and social skills squared are included as covariates in these models to adjust for psychosocial behaviors at the end of 1st grade.

Table 4 shows the first analyses that only included covariates in the model. As in the 1st-grade results, social skills from the end of 1st grade were strongly related to end-of-2nd-grade reports of social skills; girls showed higher social skills in 2nd grade as compared to boys. There were no significant effects of ethnicity. The second step included school-level assessments of DAP as a predictor of changes in social skills. In this case, DAP was a significant negative predictor of 2nd-grade social skills. To quantify this change, a child in a school at the 95th percentile

TABLE 3
Omnibus Tests for the Interaction Between DAP and Ethnicity

<i>Model/Effect</i>	<i>1st Grade</i>			<i>2nd Grade</i>			<i>3rd Grade</i>		
	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>	<i>F</i>	<i>df</i>	<i>p</i>
Social Skills									
School-level DAP	1.58	(3,957)	.23	1.07	(3,1709)	.39	1.37	(3,1786)	.44
Integrated curriculum	1.15	(3,957)	.38	.50	(3,1709)	.70	1.24	(3,1786)	.35
Social/emotional emphasis	1.81	(3,957)	.16	1.37	(3,1709)	.34	1.42	(3,1786)	.29
Child-centered approaches	2.34	(3,957)	.09	.81	(3,1709)	.53	.31	(3,1786)	.27
Problem Behaviors									
School-level DAP				.69	(3,1709)	.59	2.28	(3,1786)	.11
Integrated curriculum				1.87	(3,1709)	.51	.94	(3,1786)	.44
Social/emotional emphasis				2.68	(3,1709)	.06	1.04	(3,1786)	.44
Child-centered approaches				.47	(3,1709)	.71	.54	(3,1786)	.30

Note. DAP = Developmentally Appropriate Practice. Problem Behaviors was not assessed in 1st grade. Ethnicity was coded into four categories: African American, White, Hispanic, and all others. *F* statistics are averaged across all imputations and not corrected for uncertainty; thus, these tests likely overstate any effects that are present.

of all schools in its use of DAP would be expected to be .13 standard deviations lower than a child in a school at the 5th percentile. Although this effect is significant, it is clearly very small with few practical implications. The final main effects model included classroom-level DAP as a predictor of change in social skills. None of the three classroom components of DAP significantly predicted outcomes. In fact, their inclusion in the model slightly reduced the effects of school-level DAP, which was now no longer significant. Graphical analyses found no evidence for a nonlinear relationship of DAP with 2nd-grade social skills.

Further analyses examined the interaction of 2nd-grade social skills with gender and ethnicity. These analyses found that girls in a classroom with greater social and emotional emphasis displayed significantly lower social skills. However, if all three components of DAP increased together, this effect was canceled out by positive (but not significant) effects of a classroom's integrated curriculum and use of child-centered approaches. Tests of the interaction of DAP and ethnicity in predicting 2nd-grade social skills found no differential effects of DAP across ethnic groups in omnibus tests (see Table 3).

The final set of 2nd grade analyses examined the impacts of DAP on problem behaviors at the end of 2nd grade. The covariates-only model (reported in Table 4) showed a strong linear and a curvilinear relationship between social skills at the end of 1st grade and problem behaviors at the end of 2nd grade. This model also found females to be .13 standard deviations below males on problem behaviors and Whites to be .17 standard deviations above the grand mean on problem behaviors. The second model included school-level DAP and found no relationship between DAP and covariate adjusted problem behaviors at the end of 2nd grade. The third main effects model included classroom-level DAP factor scores and again found no effects of any aspect of DAP on end-of-2nd-grade problem behaviors. Examination of residual graphs found no evidence for nonlinear effects of DAP on problem behaviors.

The final 2nd-grade problem behaviors models examined the interactions of DAP with gender and ethnicity. No significant interactions were found for gender (see Table 4) or for the omnibus tests across all ethnic groups (see Table 3).

TABLE 4
Effects of DAP on Social Skills and Problem Behaviors in 2nd Grade

<i>Model/Effect</i>	<i>Social Skills</i>				<i>Problem Behaviors</i>			
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>p</i>
<i>Covariates Only</i>								
1st-grade social skills	.742	.016	317	.00	-.580	.019	508	.00
1st-grade social skills squared	n/a				.040	.013	168	.00
Female	.109	.031	831	.00	-.134	.037	903	.00
African American	-.002	.032	101	.94	.001	.037	903	.98
White	.010	.027	401	.71	.174	.031	179	.00
Hispanic	-.017	.045	74	.70	-.057	.049	223	.24
τ_π	.013	.007			.018	.011		
τ_β	.006	.011			.039	.017		
σ^2	.422	.016			.577	.022		
<i>School DAP</i>								
School-level DAP	-.033	.016	222	.04	.027	.021	207	.19
τ_π	.012	.007			.018	.011		
τ_β	.006	.011			.037	.017		
σ^2	.421	.016			.578	.022		
<i>Classroom and School DAP</i>								
School-level DAP	-.029	.017	208	.08	.024	.021	203	.24
Integrated curriculum	-.062	.075	257	.41	.043	.091	390	.64
Social/emotional emphasis	.108	.067	341	.11	-.124	.083	677	.14
Child-centered approaches	-.051	.060	382	.40	.072	.075	525	.34
τ_π	.011	.007			.017	.011		
τ_β	.007	.011			.039	.017		
σ^2	.421	.016			.577	.022		
<i>Interaction of DAP with Gender</i>								
Female	.114	.031	711	.00	-.137	.037	796	.00
× School-level DAP	.004	.028	412	.88	.001	.034	319	.97
× Integrated curriculum	.263	.140	386	.06	-.029	.165	564	.86
× Social/emotional emphasis	-.340	.139	133	.02	.168	.155	423	.28
× Child-centered approaches	.088	.114	437	.44	-.147	.132	942	.27
τ_π	.011	.007			.016	.011		
τ_β	.007	.011			.039	.017		
σ^2	.421	.016			.578	.022		

Note. DAP = Developmentally Appropriate Practice. Degrees of freedom differ because they are estimated *df* from multiple imputation, τ_π is the school-level variance of the intercept, τ_β is the classroom-level variance, and σ_2 is the student-level residual variance.

Effects of DAP in 3rd Grade

Further analyses examined the impacts of DAP on social skills and problem behaviors in 3rd grade. These analyses replicated those for 2nd grade, except that for the problem behaviors outcome, 2nd-grade problem behaviors were used as the covariate.

The covariates-only model for predicting social skills found a strong relationship between social skills at the end of 2nd grade and those at the end of 3rd grade. Also, African American students showed lower levels of social skills than the grand average (see Table 5). Gender and

TABLE 5
Effects of DAP Social Skills and Problem Behaviors in 3rd Grade

<i>Model/Effect</i>	<i>Social Skills</i>				<i>Problem Behaviors</i>			
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>p</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>p</i>
Covariates Only								
2nd-grade baseline ^a	.716	.018	140	.00	.745	.015	1033	.00
Female	.035	.031	1030	.26	-.041	.030	1071	.17
African American	-.064	.029	537	.03	.337	.285	178	.24
White	-.042	.027	437	.13	.042	.027	231	.11
Hispanic	.114	.413	392	.78	-.093	.041	154	.03
τ_π	.009	.007			.002	.006		
τ_β	.009	.013			.017	.010		
σ^2	.458	.017			.411	.016		
School DAP								
School-level DAP	-.012	.017	129	.47	.004	.015	148	.78
τ_π	.009	.007			.002	.006		
τ_β	.008	.013			.017	.010		
σ^2	.458	.017			.411	.016		
Classroom and School DAP								
School-level DAP	-.014	.017	139	.41	.007	.015	148	.62
Integrated Curriculum	-.040	.097	584	.68	.022	.095	387	.82
Social/Emotional Emphasis	.035	.086	477	.68	.000	.083	449	1.00
Child-centered Approaches	.019	.059	499	.75	-.043	.056	589	.45
τ_π	.009	.007			.002	.006		
τ_β	.009	.013			.017	.011		
σ^2	.458	.017			.411	.016		
Interaction of DAP With Gender								
Female	.037	.031	1014	.24	-.042	.030	1072	.16
× School-level DAP	.000	.028	544	.99	-.042	.026	759	.11
× Integrated curriculum	.150	.194	373	.44	-.031	.185	345	.87
× Social/emotional emphasis	-.103	.172	310	.55	-.037	.167	216	.83
× Child-centered approaches	-.061	.115	446	.60	.104	.107	883	.33
τ_π	.009	.007			.002	.006		
τ_β	.009	.013			.016	.011		
σ^2	.458	.017			.412	.016		

Note. DAP = Developmentally Appropriate Practice. Degrees of freedom differ because they are estimated *df* from multiple imputation, τ_π is the school-level variance of the intercept, τ_β is the classroom-level variance, and σ^2 is the student-level residual variance.

^aEnd-of-2nd-grade social skills were used as the covariate for the social skills outcome and problem behaviors for that outcome.

the other ethnic groups were not significantly related to 3rd-grade social skills. The next model included differences between schools in DAP as a predictor, and it was found not to be related to changes in social skills. The final main effects model included the three classroom-level factor scores as predictors; again, none of them was found to be related to 3rd-grade social skills. Graphical analyses showed no evidence of a curvilinear relationship between DAP and 3rd-grade social skills. Further analyses examining the interactions of DAP with gender and ethnicity also found no interactions with gender (see Table 5) and no differences in the effects of DAP across all ethnic groups (see the omnibus model tests in Table 3).

Other analyses examined the effects of DAP on 3rd-grade problem behaviors. The covariates-only model found strong effects of 2nd-grade problem behaviors and that Hispanic students were .09 standard deviations below the average student. Again, in the school-level and classroom-level models, no significant effects of DAP were observed (see Table 5). Graphical analyses also showed no evidence of a curvilinear relationship between DAP and problem behaviors. Similar to earlier analyses, when the interactions of DAP with gender and ethnicity were examined, there was no evidence of differential effects of DAP for males or females (see Table 5) or overall across the four ethnic groups examined (see Table 3).

Power

Because the main findings of this study are that classroom DAP was not consistently related to parent ratings of social skills or problem behaviors, this section reports on the ability of this study to find meaningfully large effects, had they been present. The challenge in power analyses is to calculate the standard errors for the effects of interest—in this case, the effects of DAP on social skills and problem behaviors. Here, we use the observed standard errors from Tables 2, 4, and 5 to make estimates of the effect size detectable with 80% power in these analyses. To make this more concrete, we report these effect sizes as the differences in social skills and problem behaviors that we would expect to see if a child were moved from a classroom or school at the 31st percentile in DAP to one at the 69th percentile (an increase in DAP of 1 standard deviation). For these power analyses, we assume the lowest observed degrees of freedom so that these estimates will be conservative. What we see is that at the school level, this study has 80% power to detect a true effect of DAP of .05 in the most optimistic case, to .09 in the most pessimistic case. At the classroom level, the study has 80% power to detect effects of DAP of .15 in the most optimistic case, to .27 in the most pessimistic case. These findings can be interpreted in terms of a Cohen's *d* for the effects of moving a child from a moderately low to a moderately high classroom or school. These effects are all quite small, indicating that this study can detect even quite small effects of DAP, had they existed.

DISCUSSION

The aim of this study was to better understand the effects of classroom implementation of DAP on parent perceptions of children's social skills in 1st through 3rd grades. Despite some promising findings, previous research on the impacts of DAP on psychosocial outcomes is quite limited in its scope and in its methodological rigor. This study expanded previous work by including older children (which is recommended by the DAP guidelines) and by using a much larger and more nationally representative data set than previous studies. Also, the current study used analyses appropriate to the multilevel data structure that is inherent in most work examining classroom or school effects on students. Results indicated that attending a school in which teachers' practices are more DAP oriented, or being in a classroom that emphasizes DAP, was not related to better parent ratings of social skills or problem behaviors. In fact, some evidence suggested a negative impact of schools with higher levels of classroom DAP in 2nd grade, as well as some evidence of negative effects of more social/emotionally focused classrooms for girls. However, the size of

both of those effects and the fact that they were not found consistently across years and different outcomes suggests that the emphasis should be on the conclusion that these results find no effects of classroom implementation of DAP.

Combined with previous results finding no effects of classroom aspects of DAP on academic achievement or receptive language in 1st through 3rd grades (Van Horn & Ramey, 2003), we believe that these studies make an argument for finding new directions for improving the effectiveness of education for young children. To date, no consistent evidence exists that these practices achieve their intended goal. This conclusion is supported by a critical review of the literature as well that finds that if appropriate methodology had been used, only a handful of previously published studies finding positive effects of DAP would have remained (Van Horn et al., 2005). Most of those studies that did show robust effects showed effects when children were observed in classrooms or were based on teacher reports. The current study was unable to assess whether it is the case that children in more DAP classrooms behave differently when in the classroom; however, these findings do indicate that there seem to be no effects that translate over into different settings.

This study has several limitations. Although the SSRS is one of the most widely used and accepted measures of child social outcomes in schools, we did not measure a broader array of psychosocial outcomes, such as stress, which might be affected by the use of DAP. The assessment of DAP was itself limited to those aspects that are easily observable in classroom observations; other aspects, such as assessment of performance, relationships with parents, and school policies, were not assessed and statements cannot be made about them. Further, an experimental design would provide stronger evidence for any causal effects of DAP. However, without those effects and given that the classrooms in this study differed greatly in the degree to which they conformed to the DAP guidelines, it would seem unlikely that an experimental design would lead to different results unless there were broader curriculum differences associated with the use of DAP. Also, this study did not assess the cumulative impact of being in classrooms or schools high in DAP over a longer period of time. It is possible that these impacts would be greater, although given the absolute lack of any positive effect and the high power of this study, this seems unlikely. Perhaps most important, this study does not assess the impacts of DAP on children below the 1st grade. It may be the case that DAP is appropriate for kindergarten and preschool children but becomes less effective as children pass into 1st grade.

A clear issue in the interpretation of these results is the use of parent report of social skills. Other research has shown that ratings of social skills from different sources tend to be only moderately correlated (Fagan & Fantuzzo, 1999). Although we have argued above that there is good rationale for using parent reports, it is fair to ask whether it is too high of a standard to expect that classroom practices could affect social skills observed by parents mostly outside of the classroom. This is a legitimate question, and it seems reasonable to conduct further large-scale research using independent observers.

Where does this leave us? Thus far, evidence for the use of DAP in isolation seems to be of limited value. One option is to look beyond general guidelines and focus on the broader educational context. For example, future studies should investigate how aspects of schools and classrooms and the student body come together to create an environment that allows children to achieve their educational objectives. Research that focuses on average effects or group differences may be too limited to capture the complexity of schools and their impact on children. An alternative direction would be to take the assessment of the practices suggested by the DAP to

the microlevel, examining specific practices and their impacts on student outcomes. This could lead to a refinement of the guidelines, one that focuses on those aspects of “developmentally appropriate practices” that are mostly likely to cause positive effects.

Finally, results from this study suggest that educators must understand the methodological and statistical issues that are common to many studies in education. If nested data structures are ignored, the results will have an inflated Type I error rate and other researchers will be unable to replicate the results. Indeed, numerous studies on the effects of DAP report equivocal results (many nonsignificant findings, some positive findings, and some negative findings) that often fail to be replicated. Results from this study emphasize that when studies fail to account for a nested data structure and report equivocal findings, it is quite possible that there are no true effects.

NOTE

1. The ADAPT was administered only in the last 3 years of the study, yielding the sample that includes 1st-grade classrooms with cohort II children and 2nd and 3rd grade with both cohorts. Further, as this project was funded by separate contracts for each site and then a coordinating contract with the national assessment team, each site was able to choose whether to administer the ADAPT. In all, 25 of the 31 sites used the ADAPT, although not all did so in each year.

REFERENCES

- Abbott-Shim, M., & Sibley, A. (1992). *Assessment profile for early childhood programs: Research version*. Atlanta, GA: Quality Assist.
- Abbott-Shim, M. S., Sibley, A., & Neel, J. (1998). *Psychometric report of the Assessment Profile for Early Childhood Programs: Research version of the National Head Start Demonstration Project*. Atlanta, GA: Quality Assist.
- Aldridge, J. (1992). Issues in developmentally appropriate practice and individual differences. *Journal of Instructional Psychology, 19*(2), 71–78.
- Bredenkamp, S. (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8: Expanded edition*. Washington, DC: National Association for the Education of Young Children.
- Bredenkamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs* (Rev. ed.). Washington, DC: National Association for the Education of Young Children.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models* (Vol. 1). London, England: Sage.
- Buchanan, T. K., Burts, D. C., Bidner, J., White, V. F., & Charlesworth, R. (1998). Predictors of the developmentally appropriateness of the beliefs and practices of first, second, and third grade teachers. *Early Childhood Research Quarterly, 13*(3), 459–483.
- Burts, D. C., Hart, C. H., Charlesworth, R., DeWolf, D. M., Ray, J., Manuel, K., . . . Flegge, P. O. (1993). Developmental appropriateness of kindergarten programs and academic outcomes in first grade. *Journal of Research in Childhood Education, 8*(1), 23–31.
- Burts, D. C., Hart, C. H., Charlesworth, R., Fleege, P. O., Mosley, J., & Thomasson, R. H. (1992). Observed activities and stress behaviors of children in developmentally appropriate and inappropriate kindergarten classrooms. *Early Childhood Research Quarterly, 7*(2), 297–318.
- Burts, D. C., Hart, C. H., Charlesworth, R., & Kirk, L. (1990). A comparison of frequencies of stress behaviors observed in kindergarten children in classrooms with developmentally appropriate versus developmentally inappropriate instructional practices. *Early Childhood Research Quarterly, 5*(3), 407–423.
- Charlesworth, R. (1998a). Developmentally appropriate practice is for everyone. *Childhood Education, 74*, 274–282.
- Charlesworth, R. (1998b). Response to Sally Lubeck’s “Is Developmentally Appropriate Practice for Everyone?” *Childhood Education, 74*, 293–298.
- Coie, J. D., Dodge, K. A., & Kupersmidt, J. B. (1990). Peer group behavior and social status. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 17–59). New York, NY: Cambridge University Press.

- Copple, C., & Bredekamp, S. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.
- Elkind, D. (1986). In defense of early childhood education. *Principal*, 65(5), 6–9.
- Elkind, D. (1987). Early childhood education on its own terms. In S. L. Kagan & E. F. Zigler (Eds.), *Early schooling: The national debate* (pp. 98–115). New Haven, CT: Yale University Press.
- Fagan, J., & Fantuzzo, J. W. (1999). Multirater congruence on the Social Skills Rating System: Mother, father, and teacher assessments of urban Head Start children's social competencies. *Early Childhood Research Quarterly*, 14(2), 229–242.
- Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In P. Mussen (Ed.), *The handbook of child psychology* (pp. 167–230). New York, NY: Wiley.
- Goldstein, H. (1999). *Multilevel statistical models: Kendall's library of statistics 3* (Internet ed.). London, England: Arnold Publishers.
- Gottlieb, M. (1995). *A developmentally appropriate practice template*. Des Plaines, IL: Illinois Resource Center.
- Gottlieb, M. (1997). *A developmentally appropriate practice template: Administration and technical manual*. Des Plaines, IL: Illinois Resource Center/OER Associates.
- Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Research methods in psychology* (pp. 87–114). New York, NY: John Wiley & Sons. (From *Handbook of psychology*, vol. 2, by I. B. Weiner, Ed.)
- Gresham, F. M., & Elliott, S. N. (1988). Teachers' social validity ratings of social skills: Comparisons between mildly handicapped and nonhandicapped children. *Journal of Psychoeducational Assessment*, 6, 225–234.
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Hart, C. H., Burts, D. C., Durland, M. A., Charlesworth, R., DeWulf, M., & Fleege, P. O. (1998). Stress behaviors and activity type participation of preschoolers in more and less developmentally appropriate classrooms: SES and sex differences. *Journal of Research in Childhood Education*, 12, 176–196.
- Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum.
- Hirsh-Pasek, K., Hyson, M. C., & Rescorla, L. (1990). Academic environments in preschool: Do they pressure or challenge young children? *Early Education and Development*, 1, 401–423.
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam, The Netherlands: TT-Publikaties.
- Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (Vol. 2000, pp. 147–154). New York, NY: Springer Verlag.
- Hsue, Y., & Aldridge, J. (1995). Developmentally appropriate and traditional Taiwanese culture. *Journal of Instructional Psychology*, 22(4), 320–323.
- Huffman, L. R., & Speer, P. W. (2000). Academic performance among at-risk children: The role of developmentally appropriate practices. *Early Childhood Research Quarterly*, 15(2), 167–184.
- Hyson, M. C., Hirsh-Pasek, K., & Rescorla, L. (1990). The Classroom Practices Inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly*, 5(4), 475–494.
- Jambunathan, S., Burts, D. C., & Pierce, S. H. (1999). Developmentally appropriate practices as predictors of self-competence among preschoolers. *Journal of Research in Childhood Education*, 13, 167–174.
- Jones, I., & Gullo, D. (1999). Differential social and academic effects of developmentally appropriate practices and beliefs. *Journal of Research in Childhood Education*, 14, 26–35.
- Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development*, 74(5), 1344–1368.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.
- Lubeck, S. (1998a). Is DAP for everyone? A response. *Childhood Education*, 74, 299–301.
- Lubeck, S. (1998b). Is developmentally appropriate practice for everyone? *Childhood Education*, 74, 283–292.
- Marcon, R. A. (1992). Differential effects of three preschool models on inner-city 4-year-olds. *Early Childhood Research Quarterly*, 7(4), 517–530.
- Marcon, R. A. (1993). Socioemotional versus academic emphasis: Impact on kindergartners' development and achievement. *Early Child Development and Care*, 96, 81–91.
- Marcon, R. A. (1999). Differential impact of preschool models on developmental and early learning of inner-city children: A three-cohort study. *Developmental Psychology*, 35, 358–375.

- Murray, D. M. (1998). *Design and analysis of group-randomized trials* (Vol. 27). New York, NY: Oxford University Press.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus* (Version 3). Los Angeles, CA: Muthén & Muthén.
- Ramey, S. L., Ramey, C. T., Phillips, M. M., Lanzi, R. G., Brezausek, C., Katholi, C. R., . . . Snyder, S. W. (2001). *Head Start children's entry into public school: A report on the National Head Start/Public School Early Childhood Transition demonstration study*. Washington, DC: Department of Health and Human Services, Administration on Children, Youth, and Families.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Roff, M., Sells, S. B., & Golden, M. M. (1972). *Social adjustment and personality development in children*. Minneapolis, MN: University of Minnesota Press.
- Ruckman, A. Y., Burts, D. C., & Pierce, S. H. (1999). Observed stress behaviors of 1st-grade children participation in more and less developmentally appropriate activities in a computer-based literacy laboratory. *Journal of Research in Childhood Education, 14*, 36–47.
- SAS Institute Inc. (2001a). *SAS / STAT software: Changes and enhancements* (Release 8.2). Cary, NC: Author.
- SAS Institute Inc. (2001b). *The SAS system for Windows* (Version 8.02). Cary, NC: Author.
- Schafer, J. L. (1997a). *Analysis of incomplete multivariate data*. New York, NY: John Wiley & Sons.
- Schafer, J. L. (1997b, February). *Imputation of missing covariates under a multivariate linear mixed model*. Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
- Schafer, J. L. (1997c). *NORM* (Version 2.03). State College, PA: Methodology Center at Penn State.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*(4), 323–355.
- Smith, A. B. (1996). The early childhood curriculum from a sociocultural perspective. *Early Child Development and Care, 115*, 51–64.
- Stipek, D. J., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development, 66*(1), 209–223.
- Stipek, D. J., Feiler, R., Patricia, B., Ryan, R., Milburn, S., & Salmon, J. M. (1998). Good beginnings: What difference does the program make in preparing young children for school? *Journal of Applied Developmental Psychology, 19*(1), 41–66.
- Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Lawrence Erlbaum.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York, NY: HarperCollins.
- Van Horn, M. L., Atkins-Burnett, S., Ramey, S. L., Snyder, S., & Karlin, E. O. (2007). Parent ratings of children's social skills: Longitudinal psychometric analyses of the social skills rating system. *School Psychology Quarterly, 22*, 162–199.
- Van Horn, M. L., Karlin, E. O., Ramey, S. L., Aldridge, J., & Snyder, S. W. (2005). Effects of developmentally appropriate practices on children's development: A review of research and discussion of methodological and analytic issues. *Elementary School Journal, 105*, 325–352.
- Van Horn, M. L., & Ramey, S. L. (2003). The effects of developmentally appropriate practices on academic outcomes among former Head Start students and classmates from first through third grades. *American Educational Research Journal, 40*, 961–990.
- Van Horn, M. L., & Ramey, S. L. (2004). A new measure for assessing developmentally appropriate practices: A developmentally appropriate practice template. *Early Childhood Research Quarterly, 19*, 569–587.
- Walsh, D. J. (1991). Extending the discourse on developmental appropriateness. *Early Education and Development, 2*(2), 109–119.
- Woodward, L. J., & Fergusson, D. M. (2000). Childhood peer relationship problems and later risks of educational underachievement and unemployment. *Journal of Child Psychology and Psychiatry, 41*(2), 191–201.
- Yuan, Y. C. (2000). *Multiple imputation for missing data: Concepts and new development* (No. P267-25). Cary, NC: SAS Institute.