

# Four-Year Follow-Up of Children in the LEAP Randomized Trial: Some Planned and Accidental Findings

Topics in Early Childhood Special Education  
2017, Vol. 37(2) 121–126  
© Hammill Institute on Disabilities 2017  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0271121417711531  
tecse.sagepub.com  


Phillip S. Strain, PhD<sup>1</sup>

## Abstract

This article reports on a 4-year follow-up study from the Learning Experiences and Alternative Program for Preschoolers and Their Parents (LEAP) randomized trial of early intervention for young children with autism. Overall, participants from LEAP classes were marginally superior to comparison class children on elementary school outcomes specific to communication, adaptive behavior, social, academic, and cognitive skills. Statistically significant group differences were noted in cognitive development and social skills. However, when placement was treated as an independent variable, very large effects were seen across all outcome measures, including autism symptoms, for children who were enrolled in inclusive settings. Data from adult family members confirmed important changes in perceived quality of life.

## Keywords

autism spectrum disorder (ASD), disability populations, early intervention, follow-up assessment, inclusion, evidence-based practices

“Mama always said ‘Life is like a box of chocolates, you never know what you’re gonna get.’” And thus, Forrest Gump explains his circuitous, unpredictable, and unimaginable life of twists and turns. In many ways, this article reflects a similar level of serendipity in the conduct of research. Most often, doing research in early childhood special education (ECSE), not unlike any other field of endeavor, follows a predetermined path. On occasion, however, new evidence is too tantalizing to ignore and alternative paths, questions, and insights emerge. As I began to describe this process and the resulting data in the extant case, it became abundantly clear that the “new evidence” called for its own unique presentation format. I greatly appreciate *Topics in Early Childhood Special Education* editor Erin Barton’s willingness to entertain such a divergence from the status quo of research descriptions.

## The Context for This Accidental Study

Several years ago, my colleagues and I conducted a randomized trial of the Learning Experiences and Alternative Program for Preschoolers and Their Parents (LEAP) inclusion model of early autism intervention (Strain & Bovey, 2011). In this study, we randomly assigned inclusive preschool classes (28 sites, 177 children) to receive coaching to fidelity in LEAP implementation or to receive training materials only (23 sites, 117 children). After 2 years, moderate to large effect size differences were found in favor of

children in full replication sites. Specifically, these children showed significantly better scores than comparison class children on the *Childhood Autism Rating Scale* (Schopler, Reichler, & Renner, 1988), the *Preschool Language Scale–4* (Zimmerman, Steiner, & Pond, 2001), the *Mullen Scales of Early Learning* (Mullen, 1995), and on the positive and negative behavioral dimensions of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990).

Based on these very favorable outcomes in favor of children exposed to high fidelity LEAP practices, we received funding to conduct a 4-year follow-up study to determine how these initial study group differences maintained, or not, across a 4-year period. In brief, here are the four a priori questions we addressed and their associated findings.

## What Is the Stability of Classroom Placement Across 4 Years (K–3)?

One of the more interesting data points from this follow-up study is the absolute consistency in placements for individual children across time. We observed no examples of children moving from an “autism” labeled setting to an inclusive

<sup>1</sup>University of Colorado Denver, USA

### Corresponding Author:

Phillip S. Strain, University of Colorado Denver, P.O. Box 173364,  
Campus Box 106, Denver, CO 80217-3364, USA.  
Email: phil.strain@ucdenver.edu

**Table 1.** Distribution of All Study Participants Segregated by Quartile Ranking on Preschool Outcomes as Percentage of Those in Inclusive Versus Autism Classes.

Quartile	Inclusive (%)	Autism only (%)
Fourth lowest 25%	28	23
Third 26%–50%	26	27
Second 51%–75%	28	29
First top 75%–100%	18	21
	Sum to 100% enrolled	Sum to 100% enrolled

environment nor did we see any examples of children being placed in a more restrictive setting once they were originally in an inclusive kindergarten program. Interestingly enough, a majority of sites that enrolled study graduates from both study arms in “autism” classes at kindergarten had adopted brand-named curricula and instructional practices that aimed to prepare children for inclusive settings. At least through third grade, this hoped for outcome was not observed. Based on prior research on K–12 placements, the stability of these results should not be surprising (McNulty, Widerstrom, Goodwin, & Campbell, 1988). Moreover, the factor that apparently controlled many initial kindergarten decisions, namely, the district’s unilateral policy regarding an appropriate setting for children on the autism spectrum, is consistent with prior analyses of placement decisions in ECSE classes in Pennsylvania (Miller, Boyd, Hunsicker, McKinley, Strain, & Wu, 1992).

### What Is Driving Initial Kindergarten Placement Decisions?

We found what must be considered a disturbing pattern of “child-independent” decisions for individuals in both arms of the study. Simply put, where children were placed was driven by a district-level decision conditioned on opinions about “other” individuals with autism spectrum disorder (ASD) and/or historical information. Operational policies varied widely. For example, many districts placed children from both arms of the study in “autism” classes because those children had, at least at one point, that label. Other districts essentially argued that children had made progress in an inclusive preschool and therefore they should be enrolled in an inclusive kindergarten. On occasion, aggressive action by parents altered these policy positions.

So what about children’s developmental level in placement decisions? The answer to this question is addressed in Table 1 below. Here, we show the distribution of all study participants segmented by quartile ranking on preschool

outcomes as a percentage of those in inclusive versus autism classes.

One might expect that children in the Top Quartile for outcome at preschool would represent the largest percentage of children in inclusive settings. As it turns out, they are least represented! Overall, the distribution of children in Table 1 demonstrates no correspondence between relative growth on study outcome measures in preschool and kindergarten placements.

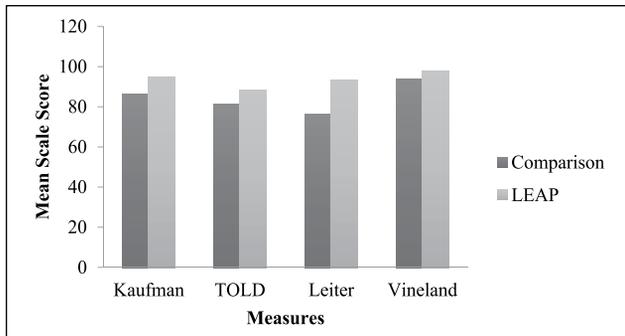
### How Did Classroom Quality Vary Across Settings?

We determined from the outset to define an inclusive placement as one in which students were in classes with typical peers 80% of the time or more. All other placements represented a residual, or less than a fully inclusive classroom. In using our quality of classroom observational measure, we found only two item categories that discriminated between groups of settings as defined (measure available from first author). Individuals using the classroom quality observation system were trained to reliability (80% or better agreement with a “gold standard” observer) prior to any data collection, and agreement percentages across observers and settings exceeded 80% for the following reported observations.

Not surprisingly, one of these items was the “Membership” scale that examined whether or not typical peers were physically in proximity when children with ASD received instruction. A “1” on the 5-point scale represented *no typical peers*, “3” represented *some peers*, and “5” represented *full peer participation* in all instruction for target children. Operationally, one might consider this scale as ranging from 1 (*tutoring for the target child in a corner of the class*) to 5 (*large group instruction for all children*).

In fully inclusive classes, the mean “membership” score was 4.7, with a range of 3 to 5. By contrast, the mean “membership” score in less than fully inclusive classes was 1.9 with a range of 1 to 4. Using a two-tailed *t* test, these mean differences are significant ( $t[90] = 2.66, p < .01$ ). In a very real sense, the stark difference in membership scores validates our a priori determination to require an 80% threshold for the inclusive designation. From an instructional standpoint, it is significant to note that the strong academic performance of children in inclusive settings took place largely without any tutorial or isolated instruction for these children.

The other item on our classroom quality scale that differentiated inclusive from noninclusive settings was the 5-point rating of classroom climate. Here a “1” represented extremely negative, unorganized, children not engaged, teachers using negative comments; “3” represented acceptable climate, children engaged 50% or more, teachers using mostly positive language; and “5” represented outstanding climate, highly engaged children, teachers consistently making positive comments, and instruction ongoing.



**Figure 1.** Mean scale scores for LEAP and comparison group participants 4 years post.

Note. LEAP = Learning Experiences and Alternative Program for Preschoolers and Their Parents; TOLD = Test of Language Development.

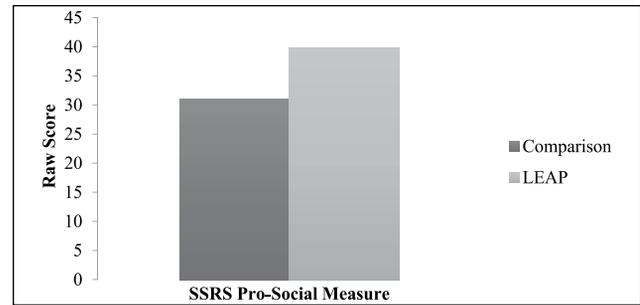
Overall, most settings scored in an acceptable range on this measure. However, statistically significant differences were noted in favor of fully inclusive settings. Their mean classroom climate rating was 4.3 with a range of 3 to 5. For “autism” classes, the mean was 3.1 and the range 2 to 4,  $t(90) = 2.01, p < .05$ . This finding is in contrast to a general expectation that children with autism are placed in less than inclusive settings because they require a more structured and organized instructional environment with higher levels of intensity, individualized instruction, and student feedback (Volkmar, Chawarska, & Klin, 2005).

### What Do Children in the LEAP Randomized Controlled Trial (RCT) Look Like 4 Years Away From Intervention?

To examine this question, we assessed original RCT participants in both study arms using a battery of measures, including the following:

- Kaufman Test of Educational Achievement, Third Edition (Kaufman & Kaufman, 2014).
- Test of Language Development–4 (TOLD; Newcomer & Hammill, 2008).
- Childhood Autism Rating System (Schopler et al., 1988).
- Leiter Brief IQ Test (Roid & Miller, 2004).
- Vineland Adaptive Behavior Scales (Sparrow, Cicchetti, & Balla, 2005).
- SSRS (Prosocial Behavior; Gresham & Elliott, 1990).

From 1 to 4 years of this follow-up study, we experienced an overall attrition rate of 32%, evenly divided between children from LEAP and comparison preschool classes. Original study participants included 177 children in LEAP classes and 117 children in comparison classes. All measures were administered in strict adherence to individual testing



**Figure 2.** Mean scores across study groups at 4 years post on the SSRS prosocial measure.

Note. LEAP = Learning Experiences and Alternative Program for Preschoolers and Their Parents; SSRS = Social Skills Rating System.

procedures and assessors had no knowledge of children’s prior study group membership.

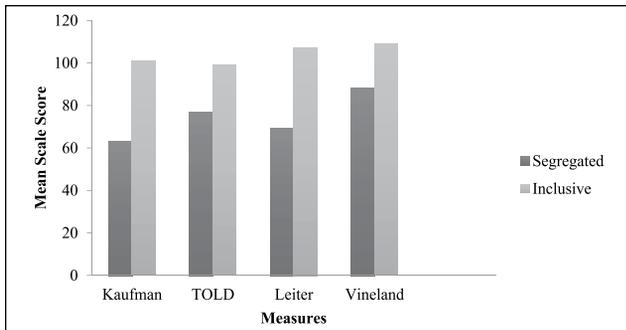
Mean standard scores for both groups at 4 years post, where 100 equals either age-level (TOLD, Leiter, Vineland) or grade-level performance (Kaufman), are illustrated in Figure 1 below.

Mean group differences on the Kaufman of 86 versus 95 in favor of LEAP graduates were not statistically significant using a two-tailed  $t$  test. Similarly, mean TOLD differences of 81 versus 88 in favor of LEAP graduates were not statistically significant using a two-tailed  $t$  test. Leiter mean differences favoring LEAP graduates of 76 versus 93 were statistically significant,  $t(177) = 2.12, p = .05$ . Using Cohen’s  $d$ , effect size for this finding was .42. Finally, mean group differences on the Vineland of 94 for comparison children and 98 for LEAP graduates were not statistically significant.

Figure 2 below shows the mean score across study groups at 4 years post on the SSRS prosocial measure. Here, higher scores represent greater perceived social skills by teachers. Differences favoring LEAP participants (31 vs. 39.8) were statistically significant,  $t(180) = 2.99, p < .01$ . Again, using Cohen’s  $d$ , the effect size for this finding was .52.

### A Data Serendipity and Overheard Conversation

Overall, these data indicate that *both* groups of children were doing well 4 years away from intervention with all measures favoring LEAP graduates. In the course of this analysis, we came to notice a very interesting trend in the data as it relates to class placement. We noticed that children with very similar preschool *Childhood Autism Rating Scale* (CARS) scores at the end of preschool were often in different placements at Kindergarten and their subsequent scores on the CARS and other measures were profoundly different 4 years post. This data trend was further heightened by my accidental eavesdropping on a conversation between data collectors outside my office. The thrust of this conversation was that data collectors were puzzled as to



**Figure 3.** Mean scale scores for CARS-matched pairs from segregated and inclusive settings 4 years post.

Note. TOLD = Test of Language Development; CARS = Childhood Autism Rating Scale.

why children scoring in the typical range of development in preschool were now in less than fully inclusive classes.

To examine this trend, we went into the database and collected as many pairs of children, regardless of study group, who left preschool with CARS scores within 3 points of each other and where one member of the pair was subsequently placed in a kindergarten with 80% or greater inclusion and the other member in a less inclusive option. We were able to detect 25 such pairs. Our rationale for creating pairs based on the CARS is that severity of autism symptoms is generally considered to be the key to determining appropriate placement in more inclusive environments for children with autism (Harris & Handleman, 1994, 2000).

For each measure in Figure 3, large differences that were statistically significant at  $p < .01$  using two-tailed  $t$  tests were observed: Kaufman,  $t(20) = 2.92$ ; TOLD,  $t(20) = 2.96$ ; Leiter,  $t(20) = 3.11$ ; Vineland  $t(20) = 2.99$ . The effect sizes for these findings were Kaufman  $d = .72$ , TOLD  $d = .66$ , Leiter  $d = .71$ , and Vineland  $d = .81$ . Figure 3 graphically displays the differences.

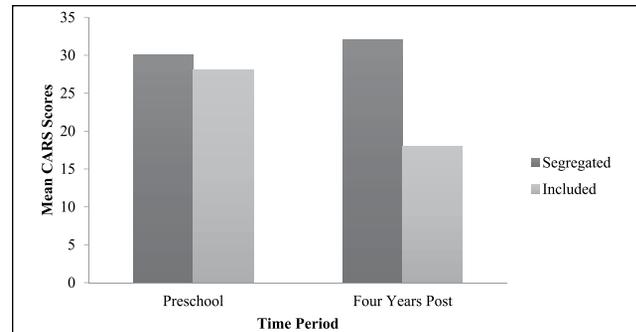
Figure 4 below shows the mean CARS scores at end of preschool and at 4 years post for these same segregated and inclusive pairs.

The mean CARS score difference post 4 years was significantly different in favor of the included member of the pairs, two-tailed  $t(20) = 2.99$ ,  $p < .01$ . The effect size for this finding was .68.

Figure 5 shows the mean SSRS raw scores for each study group 4 years post on the Prosocial Skills subscale. The 43.4 versus 24.5 difference in favor of included children was significant,  $t(20) = 3.44$ ,  $p < .05$ . The effect size for this finding was .48.

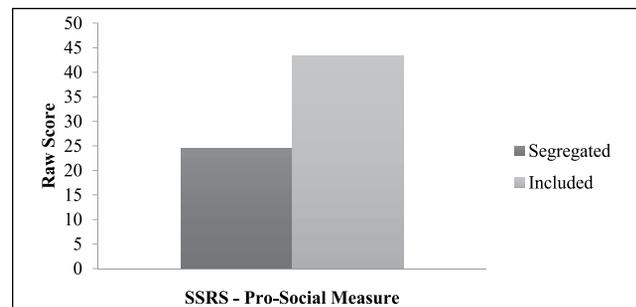
### What Might Account for These Differences?

Taken together, the results specific to placement and outcome indicate that placement might better be considered as an independent rather than a dependent variable in early intervention follow-up studies. To further explore the stark differences in



**Figure 4.** Mean CARS scores at preschool and 4 years post for “matched” CARS pairs assigned to segregated and included kindergarten.

Note. CARS = Childhood Autism Rating Scale.



**Figure 5.** Social Skills Rating System prosocial skills raw subscale scores for each study group 4 years post.

outcomes observed, we added another design variation to the planned follow-up study and did follow-up interviews with study data collectors to get some indication of what might be at play to account for these differences. From these open-ended interviews, several themes emerged when these individuals talked about other than the measured differences between placements. Standard methods for arriving at themes from these qualitative data were utilized (Dey, 1993).

The first theme centered on curriculum, a variable we did not measure directly. In the inclusive settings, the children in these classes were reported to be full participants in the regular education curriculum. By contrast, in the more segregated settings, districts had often adopted remedial curricula and “autism” curricula, or developed their own “autism” curricula. These curricula, for the most part, were not focused on age or grade-level academic content and in many ways actually mimicked content that children were exposed to at preschool (e.g., shapes, object names, colors, etc.). Put simply, data collectors uniformly reported that many children in less than fully inclusive settings were not challenged developmentally at school. Moreover, the observers could not recall that children in these settings were ever required to do work outside of school, namely, homework. In other words, children were apparently exposed to a very different curriculum at a very different dosage level.

**Table 2.** Questions and Associated Family Member Response Themes.

Question	Associated response themes
1. Why and when were your concerns about _____ developed?	1. Lack of communication at age 3 2. Obsession with objects at age 2 3. Socially isolating self at age 3
2. If you only had three words to describe _____ at the time he or she was diagnosed, what would they be?	1. Distant 2. Frustrated 3. Rigid
3. What three words would you use to now describe _____?	1. Happy 2. Kind 3. Intelligent
4. How have your expectations changed for _____ over the last 3 years?	1. Expect more in general 2. Expect better future 3. Expect more socially
5. How has family life changed over the last 3 years?	1. It hasn't 2. Do more things as a family 3. _____ has taken on more responsibilities
6. What do you see as the most important outcomes from _____'s LEAP experience in preschool?	1. Social improvement (from peers) 2. Communication improvement 3. Getting to go more places
7. In what activities at school and outside school does _____ participate on a regular bases? (All families listed at least three activities).	1. Sports 2. Music (e.g., singing, orchestra) 3. Church activities
8. Does _____ have any best friends that he or she plays with and visits with regularly?	71% indicate yes

Note. LEAP = Learning Experiences and Alternative Program for Preschoolers and Their Parents.

The second theme that emerged was related to the specific instructional supports offered by paraeducators in both types of settings. In the segregated contexts, paraeducator activities could be classified in two general ways: either (a) they were providing extremely high levels of support to children, which discouraged children's independent participation, or (b) they spent most of their time doing paperwork and housekeeping kinds of chores. Alternatively, in inclusive settings, paraeducators were regularly seen assisting children with academic assignments and particularly providing additional cues (e.g., models, partial physical prompts, prompts to peers) such that children completed tasks accurately and as independently as possible. Again, this reported behavior pattern perhaps speaks to a different "dosage" and quality of instruction.

The final theme that emerged involved the often talked about (but rarely directly measured) concept in education of high expectations. This theme was manifest as follows. In segregated settings, observers noted that teachers rarely recognized, commented on, or praised children for correct responding related to preacademic or academic tasks. In contrast, the majority of their feedback to children was related to behavior management, encouraging compliance with requests, and general task engagement. Relatedly, when children were not correct in responding, they seldom received corrective feedback. In inclusive settings, teaching staff regularly gave children feedback on their class work, grading activities, suggesting correct answers, and generally holding all children to a standard of accuracy.

### *A Quality of Life Footnote to This Accidental Study*

Given the differences between members of the twin pairs, we were curious to see how adult family members felt about their children's progress from preschool through third grade and how child quality of life was viewed. Interestingly enough, we can find no data in the autism literature where investigators asked adult family members to reflect on this topic. Adult family members of children from the 25 selected pairs who were enrolled in inclusive settings in kindergarden through third grade responded orally via the telephone to eight questions and their responses were transcribed. Following recommendations by Dey (1993), we used the key-word-in-context (KWIC) qualitative methodology to derive themes for each question, with the exception of Question 8. Listed below in Table 2 are the questions and associated themes. The themes are ordered by their relative frequency.

### **Summary and Future Research Directions**

Much of what we do in the name of research in ECSE can be considered a necessary confirmation of the obvious. Occasionally, however, pursuing the scientific methods in a usual pedestrian fashion yields a box of chocolates and the opportunity to explore the not so obvious. I would argue that this is very much in the spirit of research articulated by Sidman (1960) who talked eloquently about research studies evolving based on incoming data. Essentially, the point

is this: It is well and good to prespecify design elements and it is also well and good to permit incoming data to alter our plans, theories, and preconceived notions about what we think we know to be true.

What started out as a study of LEAP participant follow-up evolved into something perhaps more impactful, generalizable, and of far more heuristic value. These powerful differential results for children placed in inclusive elementary school environments of course need replication. They need replication with more functional, observational measures of children behaving in authentic settings. They need replication with better measures of curricular variables and follow-up dosage of instruction. They need replication with direct observational measures of teacher–child interaction. They certainly need replication across diverse groups of children.

Notwithstanding the need for replication as defined above, I would submit that the current data are sufficient to occasion an important rethinking, or at least some caution, in our field about inclusive class placement as solely an outcome index in our follow-up research. That historical perspective about placement is based on the notion that the behavior change achieved in our early childhood programs provides access to these settings. The data in this study reveal a very different picture with placement driven by district policy primarily and child progress having no obvious relationship with placement.

Finally, these data speak to what I have referred to as the necessity for longitudinal, quality inclusion to truly evaluate this instructional arrangement (Strain, 2016a, 2016b). In this regard, I would suggest that inclusion has yet to be tried, as I know of no cohort study of children who have been in quality inclusive settings *throughout* their schooling. The data from this follow-up clearly show the necessity to study such an administrative arrangement with a longitudinal lens.

Forty-six years ago, I wrote what I thought was a stunning, yet pessimistic, article about the quality of life chances for individuals with autism in a class taught by Bill Bricker at Peabody College. I got a C, a chance to rewrite (I took it), and the following comment: “Only after the perfectly designed intervention is implemented by the perfectly trained personnel can you begin to speculate about the capabilities of people with disabilities.” Perhaps this is true as well for longitudinal, quality inclusion and what it might yield.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support received from the Institute of Education Sciences, U.S. Department of Education, to the University of Colorado Denver, Grant R324A110246.

### References

- Dey, I. (1993). *Qualitative data analysis: A user guide for social scientists*. London, England: Cambridge University Press.
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Harris, S. L., & Handleman, J. S. (Eds.). (1994). *Preschool education programs for children with autism*. Austin, TX: Pro-Ed.
- Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four- to six-year follow-up. *Journal of Autism and Developmental Disorders*, *30*, 137–142.
- McNulty, B., Widerstrom, A. H., Goodwin, L. D., & Campbell, S. (1988). Who should be served, where and why: Local special education administrators' views. *Topics in Early Childhood Special Education*, *8*, 51–60.
- Miller, L. J., Boyd, K., Hunsicker, S., McKinley, J., Strain, P. S., & Wu, A. (1992). Parental attitudes toward integration. *Topics in Early Childhood Special Education*, *12*, 230–246.
- Mullen, E. M. (1995). *Mullen Scales of Early Learning: AGS edition*. Circle Pines, MN: American Guidance Service.
- Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development—Fourth edition: Primary*. Austin, TX: Pro-Ed.
- Roid, G. H., & Miller, L. J. (2004). *Leiter International Performance Scale—Revised*. Wood Dale, IL: Stoelting.
- Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *Childhood Autism Rating Scale*. Austin, TX: Pro-Ed.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.
- Strain, P. S. (2016a, August). *Some personal thoughts on special education research: Past, present and future*. Keynote presented at the Office of Special Education Project Directors Conference, Washington, DC.
- Strain, P. S. (2016b, October). *Hopes and dreams about early childhood special education research*. Keynote presented at the Division for Early Childhood Conference, Louisville, KY.
- Strain, P. S., & Bovey, E. H. (2011). Randomized, controlled trial of the LEAP model of early intervention for young children with autism spectrum disorders. *Topics in Early Childhood Special Education*, *31*, 133–154.
- Volkmar, F., Chawarska, K., & Klin, A. (2005). Autism in infancy and early childhood. *Annual Review of Psychology*, *56*, 315–336.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2001). *Preschool Language Scale* (4th ed.). San Antonio, TX: The Psychology Corporation.