J Autism Dev Disord (2014) 44:366–380
DOI 10.1007/s10803-013-1877-9

ORIGINAL PAPER

Comparative Efficacy of LEAP, TEACCH and Non-Model-Specific Special Education Programs for Preschoolers with Autism Spectrum Disorders

Brian A. Boyd · Kara Hume · Matthew T. McBee · Michael Alessandri · Anibal Gutierrez · LeAnne Johnson · Laurie Sperry · Samuel L. Odom

Published online: 28 June 2013

Abstract LEAP and TEACCH represent two comprehensive treatment models (CTMs) that have been widely used across several decades to educate young children with autism spectrum disorders. The purpose of this quasi-experimental study was to compare high fidelity LEAP ($n = 22$) and TEACCH ($n = 25$) classrooms to each other and a control condition ($n = 28$), in which teachers in high quality special education programs used non-model-specific practices. A total of 198 children were included in data analysis. Across conditions, children's performances improved over time. This study raises issues of the replication of effects for CTMs, and whether having access to a high quality special education program is as beneficial as access to a specific CTM.

Keywords Comparative effectiveness · Comprehensive treatments · Efficacy · Intervention · LEAP · TEACCH

Introduction

Providing children with autism spectrum disorders (ASD) access to high quality, early intervention results in improved developmental performance (Boyd et al. 2010;

Electronic supplementary material The online version of this article (doi:10.1007/s10803-013-1877-9) contains supplementary material, which is available to authorized users.

B. A. Boyd (✉)
Division of Occupational Science and Occupational Therapy, University of North Carolina at Chapel Hill, 2050 Bondurant Hall, Chapel Hill, NC 27599, USA
e-mail: brian_boyd@med.unc.edu

K. Hume
FPG Child Development Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: kara.hume@unc.edu

M. T. McBee
Educational and Quantitative Psychology, East Tennessee State University, Johnson City, TN, USA
e-mail: mcbeem@etsu.edu

M. Alessandri
Center for Autism and Related Disabilities, University of Miami, Coral Gables, FL, USA
e-mail: malessandri@psy.miami.edu

M. Alessandri
Department of Psychology and Pediatrics, University of Miami, Coral Gables, FL, USA

A. Gutierrez
Department of Psychology, Florida International University, University Park, FL, USA
e-mail: anibal.guiterrez@fiu.edu

L. Johnson
Special Education Programs, University of Minnesota, Minneapolis, MN, USA
e-mail: chaf0032@umn.edu

L. Sperry
Autism Centre of Excellence, School of Arts, Education and Law, Griffith University, Brisbane, Australia
e-mail: l.sperry@griffith.edu.au

S. L. Odom
FGP Child Development Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: slodom@unc.edu

S. L. Odom
School of Education, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Dawson et al. 2010; Kasari et al. 2006; Landa et al. 2011; Stahmer and Ingersoll 2004). However, debate persists over which treatment approach(es) to use to attain that ultimate goal. Presently, there are two overarching categories of intervention approaches from which practitioners or families can select treatments—focused interventions or comprehensive treatment models (CTMs). Focused interventions refer to treatments that are typically shorter in duration and target discrete skills (e.g., communication or challenging behavior). CTMs employ focused intervention practices that are organized around a central theoretical or conceptual framework, are typically used with children for a longer period of time, and target multiple developmental domains (Odom et al. 2010a).

Most of the empirical support for autism-specific, evidence-based practices comes from the focused intervention literature (Odom et al. 2003, 2010b). For instance, Odom et al. (2010a) identified 30 CTMs for individuals with ASD, and one of the resounding and surprising findings of their literature review was the lack of empirical evidence for the majority of CTMs. The most notable exception is the research related to the Lovaas Young Autism Project, with multiple studies (Eikeseth et al. 2002; Lovaas 1987; Smith et al. 2000) as well as meta-analyses (Eldevik et al. 2009; Reichow and Wolery 2009) demonstrating evidence for this CTM. However, the research on this model has been primarily conducted in clinic- or home-based settings. Like most children, those with ASD spend a great deal of their time in schools; therefore, it is of critical importance to establish the efficacy of school-based CTMs.

Increasingly, leaders in the field are calling for comparative efficacy studies to determine the relative effects of treatments when compared to each other (Dingfelder and Mandell 2011; Sox and Greenfield 2009). Two CTMs that have a long history in the field, are used frequently, and have different conceptual frameworks are LEAP (Learning Experiences and Alternative Program for Preschoolers and their Parents) and the TEACCH Autism Program. The purpose of this study was to examine the relative effects of LEAP and TEACCH school-based CTMs when compared to each other and a control condition consisting of non-model-specific (NMS) special education programs.

TEACCH and LEAP are two long-standing CTMs, their origins for individuals with ASD being traced to the 1970s and 1980s, respectively. Yet, the philosophical tenets as well as actual practices underlying these CTMs represent quite divergent approaches to educating children with ASD. TEACCH bases its conceptual orientation in cognitive-social learning theory and subscribes to a “culture of autism,” in which accommodations such as visual schedules and work systems (Hume and Odom 2007; Hume et al. 2012) are made to the environment versus the individual to promote the child’s engagement and learning (Mesibov and Shea 2010; Mesibov et al. 2005). While it is not a stated principle of TEACCH, in the context of schools, this has often manifested itself in children with ASD being educated together in classrooms that are separate from their typically developing peers. In contrast, LEAP bases its treatment approach on a blend of applied behavior analysis (ABA) as well as common tenets of early childhood education (ECE) (Strain and Hoyson 2000; Strain et al. 1985) with a goal of reducing children’s characteristics of autism that interfere with their learning opportunities. The LEAP model uses an inclusive education approach whereby children with ASD are taught alongside typically developing, same-aged peer confederates who become agents of social instruction and intervention. Thus, LEAP and TEACCH provide “naturally occurring” contrasts with the former model using an inclusive education approach and the latter primarily implemented in classroom settings emphasizing more structured, adult-led learning opportunities. Yet the evidence base for LEAP and TEACCH, at least until recently (see Strain and Bovey 2011, for research on LEAP), has been relatively sparse and there still are no existing studies that have directly compared the two approaches.

With TEACCH, some evidence exists which documents positive effects of this CTM when implemented in home (Ozonoff and Cathcart 1998), residential (Panerai et al. 1998, 2002), and school settings with older students (Panerai et al. 2009). However, we are aware of only one study in which TEACCH was evaluated as a school-based CTM for preschool-aged children with ASD (Tsang et al. 2007). In this quasi-experimental study, which was conducted in China, Tsang et al. assigned a total of 34 preschool-aged children to a TEACCH (N = 18) or a services-as-usual (N = 16) control condition. Tsang and colleagues found some improvements in favor of the experimental group in certain developmental skills (e.g., fine and gross motor) but no group differences in the areas of socialization or communication. Obviously, issues of culture as well as differences in the education system limit any generalization of the study’s findings to children with ASD served in the US public school system. In contrast, a recent large-scale efficacy study of the LEAP model demonstrated positive effects in the domains of socialization, cognition, language and challenging behavior for preschool-aged children with ASD (N = 177 LEAP, 117 = control). This study used a cluster randomized trial design with a total of 56 inclusive classrooms randomized to either the treatment condition in which teachers received 2 years of training and ongoing consultation in LEAP implementation, or a control condition in which teachers only received access to the LEAP treatment manual and training presentation materials. This first large-scale study of LEAP demonstrated that the program is efficacious as well as socially valid when fully implemented; however, a question remains as to whether it is more effective than other school-based CTMs designed for children with ASD. Comparative efficacy studies allow us to address these important causal-comparative or effectiveness research questions.

Randomized clinical trials (RCTs) represent the gold-standard in medical and education-related fields for conducting efficacy research. However, there are circumstances under which RCTs are not viable options and even reactive with certain groups (Shadish et al. 2002). For instance Strain and Bovey (2011) demonstrated, in their study, that teachers must implement LEAP for at least 2 years to find the most robust treatment effects. This would require teachers in any comparison condition to have a similar amount of training time to rule out the sheer amount of time teachers and children are exposed to the LEAP model (i.e., dosage) as being the primary explanatory variable for any observed treatment differences. In addition, research has demonstrated that when a philosophical mismatch exists between teachers' beliefs and the underlying tenets of a model that they are more inclined to experience burnout (Jennett et al. 2003). This could certainly occur in a study that randomized teachers to LEAP or TEACCH conditions given that one approach espouses an inclusive ABA/ECE approach for children with ASD, and the other involves structured learning approaches that often occur in self-contained classroom settings. Therefore, the current study used a rigorous quasi-experimental study design to compare the LEAP and TEACCH treatments because of these identified issues as well as the fact that both approaches have already been widely disseminated in public school systems throughout the US.

Specifically, a quasi-experimental study design was used to compare high fidelity and quality LEAP and TEACCH classrooms to each other as well as a control condition consisting of high quality special education classrooms in which teachers did not use practices that aligned with any particular theoretical or conceptual model. The following research questions were addressed in this study:

1. What are the effects of LEAP, TEACCH and NMS control classroom experiences on the developmental and behavioral performance of preschool-aged children with ASD?
2. What family or child factors moderate intervention effects for children in LEAP, TEACCH or NMS control classrooms?

Methods

This multi-site study was conducted in public school districts in the following states: North Carolina, Colorado, Florida and Minnesota. Research personnel from the Universities of North Carolina-Chapel Hill and Miami were involved in the collection of child data for 3 years of the project; whereas the Universities of Colorado-Denver and Minnesota were involved for two project years, primarily because of feasibility issues (e.g., the number of classrooms that could be recruited within a given site). The Institutional Review Boards at each of the respective study sites approved the research.

Sample

Classroom teachers and children in this study comprised one of three mutually exclusive groups: LEAP, TEACCH or NMS classrooms.

Classrooms/Teachers

A total of 25 TEACCH, 22 LEAP and 27 NMS teachers were enrolled in the study. Teachers were enrolled in the

Table 1 School and teacher demographics by treatment model

	TEACCH		LEAP		NMS	
	n	%	n	%	n	%
School setting						
Urban	13	52.00	10	45.45	18	64.29
Suburban	12	48.00	11	50.00	10	35.71
Rural	0	0.00	1	4.55	0	0.00
Minority enrollment	.	56.90	.	53.52	.	54.33
Teacher race						
White	24	96.00	21	95.45	27	100.00
Black	1	4.00	1	4.55	0	0.00
Teacher gender						
Male	0	0.00	1	4.55	0	0.00
Female	25	100.00	21	95.45	27	100.00
Teacher highest degree earned						
AA	0	0.00	1	4.55	1	3.70
BS/BA	9	36.00	6	27.27	13	48.15
MA/MS	16	64.00	14	63.64	11	40.74
Above MA/MS	0	0.00	1	4.55	2	7.41
Duration of school day						
1/2 day	5	20.00	22	100.00	21	75.00
Full day	20	80.00	0	0.00	7	25.00
Type (NMS only)						
Inclusive	.	.	.	.	12	42.86
Special education	.	.	.	.	16	57.14
	TEACCH		**LEAP**		**NMS**	
	Mean	SD	Mean	SD	Mean	SD
Years teaching	7.72	4.51	11.86	6.29	11.41	7.24
Adult–child ratio	2.65	1.45	4.94	1.53	4.23	2.31

One NMS teacher taught in two separate NMS classrooms study for the duration of one school year and could not participate again in subsequent years. Relevant demographic information on schools, classrooms and teachers can be found in Table 1.

Inclusion/Exclusion Criteria Teachers/classrooms were enrolled in the study if they met the following inclusion criteria: (1) all classrooms had to operate within the public school system; (2) teachers had to be certified to teach in their respective state; (3) TEACCH and LEAP teachers must have attended a formal training, either conducted by personnel directly affiliated with those programs or conducted by others who had been formally trained; (4) teachers must have been teaching in their respective classroom type for at least 2 years prior to study enrollment; and (5) teachers must have met prior-determined criteria on classroom fidelity and/or quality rating scales. Specifically, all classrooms had to meet an "average" rating (score of 3 out of 5) on four subscales of a validated classroom quality measure—the PDA Program Assessment (Professional Development in Autism Center 2008) during an initial classroom visit. In addition, TEACCH and LEAP classrooms had to meet above average ratings (3.5 out of 5) on model-specific subscales and items on their respective fidelity of implementation measures. If required average scores were not met on the initial visit, research staff conducted one additional classroom visit using the same criteria described above to make a final determination on study eligibility. Classrooms that did not meet study requirements were excluded from the study. See Fig. 1 for a CONSORT diagram describing the enrollment process.

Booster Training All eligible TEACCH and LEAP teachers subsequently participated in a booster training that occurred prior to their start in the study. The booster training was a 12-h training conducted across 2 days and the content was tailored to address any domains on the fidelity measure that received average or below average scores during the initial screening(s) (e.g., communication, behavior management). Trainers who were approved/certified by model developers conducted this training. Booster training was not offered to NMS teachers because the focus of the study was the active TEACCH and LEAP treatment models.

Children/Families

Based on participant enrollment data, a total of 205 children were initially enrolled into the study. However, seven children were excluded from data analysis because they did not meet the study diagnostic criteria outlined below. Thus, 198 children (N = 85 TEACCH, 54 LEAP, and 59 NMS) were initially included in data analysis. Demographic information on children/families can be found in Table 2.

Inclusion/Exclusion Criteria Children were enrolled in the study if they met the following criteria: (1) between 3 and 5 years of age at time of enrollment; (2) previous clinical diagnosis or educational label consistent with ASD or developmental delay; (3) met diagnostic criteria on Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999) and/or Social Communication Questionnaire (SCQ; Rutter et al. 2003); (4) had not been previously exposed to the comparison CTM, for example, a child enrolled in a TEACCH classroom could not have been previously enrolled in a LEAP classroom; and (5) must have a minimum of 6 months of exposure to the treatment or control condition. Children with significant uncorrected vision or hearing impairment, uncontrolled seizure disorder or traumatic brain injury were excluded from the study. Finally, families must have been proficient enough in English to participate in order to complete parent rating scales.

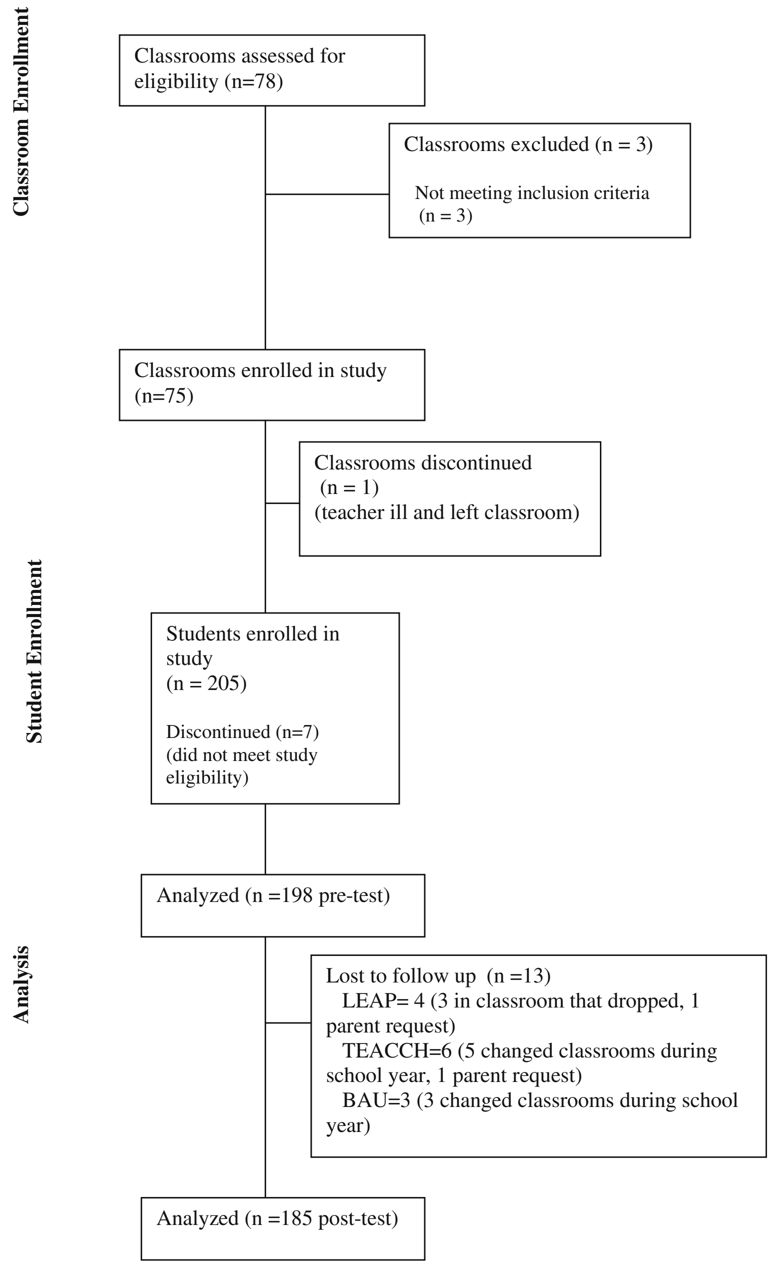
Recruitment/Retention

In general, teachers were first recruited into the study through local school administrative contacts. Administrators were asked to identify high-quality LEAP, TEACCH and NMS programs, and the observational screening protocol described above to document classroom fidelity and quality was used to confirm group assignment. The model purveyors and their staff were enlisted to identify high fidelity classrooms implementing TEACCH and LEAP; however, these classrooms also had to meet study screening inclusion criteria. Finally, all NMS classrooms were recruited from within the same school district as TEACCH and LEAP classrooms, and these classrooms were a mixture of inclusive and self-contained classrooms to minimize potential confounds. Following teacher recruitment, teachers or project staff then sent consent forms home to eligible families. Teachers were paid a total of $500 for their study participation, and families were paid a total of $200.

Data Collection

Research staff conducted the majority of direct child assessments at children's schools. On some occasions direct child assessments were conducted in clinic or home settings to collect additional measures on children. For parent report data, we mailed assessment packets to parents and conducted follow-up visits in the home. Teacher report data were collected by dropping off and picking up assessment packets at the child's school. Data collection occurred at the beginning (N = 198) and end of the school year (N = 185), at least 6 months apart. Across the three data sources (parent, teacher and child), we collected all

Fig. 1 CONSORT diagram



data on an individual child within a 6-week time window at pre and posttest timepoints.

Procedures for Collection of Fidelity Data Trained and reliable research staff collected data on the fidelity of implementation of TEACCH and LEAP practices, as well as overall classroom quality (i.e. PDA Program Assessment), four times across the school year. Training and reliability procedures as well as psychometric data on the measures are detailed in Hume et al. (2011). Two of the four visits included a reliability observer. Mean inter-rater reliability for the TEACCH measure in TEACCH classrooms was 98 % (range 88–98 %), 95 % for the LEAP measure in LEAP classrooms (range 88–100 %), and 88 % for the PDA Quality measure in NMS schools (range 87–99 %). To ensure consistency across sites, all fidelity observations occurred during the first 3 h of the classroom day. During each classroom visit and across classroom types, the two model-specific fidelity measures as well as the quality measure were completed (e.g., in a TEACCH classroom both the TEACCH and LEAP fidelity measures would have been scored in addition to the PDA quality measure). This was done to assess the degree of overlap between the classroom types. All measures were scored on a 1 (no/minimal implementation) to 5 (full implementation) scale. Based on the fidelity measures, the classrooms maintained fidelity to their respective model over time, as evidenced by the small standard deviations, however, there was also overlap across models (see Table 3 for results).

Table 2 Child and family demographics by treatment model

	TEACCH		LEAP		NMS	
	n	%	n	%	n	%
Child race						
White	32	37.6	25	46.3	35	59.3
Black	14	16.5	3	5.6	6	10.2
Hispanic	31	36.5	23	42.6	15	25.4
Asian	5	5.9	2	3.7	3	5.1
Missing	3	3.5	1	1.9	0	0.0
Child gender						
Male	71	83.5	42	77.8	52	88.1
Female	14	16.5	12	22.2	7	11.9
Missing	0	0.0	0	0.0	0	0.0
Primary caregiver gender						
Male	12	14.1	7	13.0	2	3.4
Female	72	84.7	46	85.2	53	89.8
Missing	1	1.2	1	1.9	4	6.8
Caregiver education						
Less than college	44	51.8	25	46.3	25	42.4
College or higher	39	45.9	28	51.9	32	54.2
Missing	2	2.4	1	1.9	2	3.4
Annual household income						
<$20 k–$39,999	30	35.3	14	25.9	16	27.1
$40 k–$79,999	22	25.9	18	33.3	13	22.0
>$80 k	29	34.1	18	33.3	25	42.4
Missing	4	4.7	4	7.4	5	8.5
	TEACCH		LEAP		NMS	
	Mean	SD	Mean	SD	Mean	SD
Child age at enrollment	4.00	0.57	3.96	0.70	4.07	0.64

Table 3 Mean fidelity scores and quality rating scores by treatment model

Classroom type	TEACCH		LEAP		NMS	
Type of fidelity measure	M	SD	M	SD	M	SD
TEACCH measure	4.3	0.45	3.8	0.46	3.4	0.54
LEAP measure	4.0	0.45	4.6	0.29	4.2	0.49
Quality measure	4.2	0.38	4.7	0.23	4.3	0.34

Supplemental Classroom Practices Teachers completed the *Classroom Practice Inventory* (CPI; TEACCH-LEAP project team 2007) at pretest and posttest in order to self-report the type and frequency of supplemental teaching strategies used to educate children with ASD in their classrooms. Preliminary psychometric data on the CPI indicate it is a reliable tool (ICC = 0.80; α = 0.77) that identifies commonly used classroom practices (e.g., discrete trial training, pivotal response treatment, behavioral supports). The measure is scored on a 0–4 scale, with higher scores indicating more frequent use of the practice. Descriptive data from the CPI indicated that there was similar use of some instructional strategies across model type (e.g., mean visual support scores were: NMS = 3.82, LEAP = 3.91, and TEACCH = 3.71), while other practices were more associated with specific models (e.g., mean peer-mediated instruction scores were: NMS = 1.00, LEAP = 3.18, and TEACCH = 1.17). See Table 4 for results.

Measures

The following measures were collected on children as part of their study participation: Autism Diagnostic Observational Schedule (Lord et al. 1999); Childhood Autism Rating Scale (Schopler et al. 1988); Leiter International Performance Scale-Revised (Roid and Miller 1997); Mullen Scales of Early Learning (Mullen 1995); Pictorial Infant Communication Scales (Delgado et al. 2001); Preschool Language Scales, 4th Edition (Zimmerman et al. 2002); SCQ (Rutter et al. 2003); Social Responsiveness Scale (Constantino 2002); Repetitive Behavior Scales-Revised (Bodfish et al. 1999); and Vineland Adaptive Behavior Scales, Survey Edition (Sparrow et al. 1984). Study measures included as outcomes or moderators were, in part, selected based on prior research on children with ASD (e.g., Farmer et al. 2012; Reichow and Wolery 2009), and research on the CTMs under study (e.g., Strain and Bovey 2011; Tsang et al. 2007). Descriptions of the measures and their psychometric properties can be found in the supplementary material Appendix A. Children's scores at baseline and posttest on these individual measures can be found in supplementary material Appendix B. However, baseline group differences are only reported for the composite variables (described below) because these variables served as the actual outcomes for this study.

Results

Derivation and Empirical Validation of Composite Variables

This study involved the measurement of a large number of cognitive, behavioral, psychological, and social variables, many of which could have been responsive to treatment and could have therefore served as outcome variables. However, we chose to construct composite variables for the following reasons: (a) because fitting models to a large number of correlated outcome variables would have been problematic due to the difficulty in conceptualizing patterns of results across outcomes, (b) the power-reducing effects of measurement error in the outcome variables, and

Table 4 Classroom practice inventory (CPI) scores by treatment model

CPI item/composite	NMS			LEAP			TEACCH		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Item 1. Discrete trial training	27	0.89	1.31	22	1.27	1.39	24	1.33	1.61
Item 2. Pivotal response training	27	1.63	1.71	22	0.91	1.31	24	0.88	1.33
Item 3. Floortime or DIR	27	2.89	1.37	22	2.50	1.47	24	2.42	1.06
Item 4. Relationship development intervention*	27	0.30	0.72	22	1.23	1.63	24	0.88	1.19
Item 5. Structured teaching (TEACCH)***	26	2.58	1.55	22	2.36	1.33	24	4.00	0.00
Item 6. Learning experiences—an alternative program for preschoolers and parents (LEAP)***	26	1.19	1.63	22	3.86	0.47	24	0.25	0.85
Item 7. Incidental or naturalistic teaching	26	1.96	1.69	22	2.64	1.29	24	1.88	1.45
Item 8. Picture exchange communication system*	27	1.74	1.48	22	1.68	1.52	24	2.71	1.27
Item 9. Verbal behavior	26	1.12	1.61	22	0.64	0.90	24	1.08	1.28
Item 10. Functional communication training	26	1.00	1.39	22	0.68	1.13	24	1.08	1.35
Item 11. Visual supports	27	3.82	0.62	22	3.91	0.29	24	3.71	0.86
Item 12. Video modeling	27	0.00	0.00	22	0.18	0.85	24	0.42	1.02
Item 13. Voice output communication**	27	0.52	0.75	22	1.46	1.60	24	1.58	1.44
Item 14. Computer assisted instruction	27	0.67	1.36	22	0.59	1.26	24	0.50	1.06
Item 15. Social stories/comic book conversations	27	1.59	1.25	22	2.09	1.31	24	2.00	1.25
Item 16. Peer-mediated instruction***	27	1.00	1.59	22	3.18	1.33	24	1.17	1.27
Item 17. Social skills training*	27	3.37	1.04	22	3.73	0.70	24	2.88	1.23
Item 18. Self-management strategies	27	1.15	1.59	22	0.77	1.23	24	0.75	1.11
Item 19. Positive behavior supports	27	3.48	1.05	22	3.36	0.95	24	3.21	1.10

Separate one-way ANOVAs were performed for all items. * $p < .05$; ** $p < .01$; *** $p < .001$

(c) the expected increase in the family wise error rate resulting from the estimation of a large number of parameters across the models.

In order to create the composite variables, we first collected the complete set of possible outcome variables (listed above in measures). Variables were placed into prospective composite variables based on the theoretical constructs measured as well as an examination of simple correlation coefficients. In the next step, we used exploratory factor analysis (EFA) to determine whether the conceptual groupings from the first stage were supported by the data. We chose an oblique rotation for this step. Based on the examination of factor loadings and scree plots, we revised the conceptual groupings. For example, we originally conceptualized separate expressive and receptive communication factors, but the EFA results strongly suggested that they should be combined into a single communication factor. In the last step, we used confirmatory factor analysis (CFA) to provide a final check of the construct validity of the composite variables and to estimate factor scores to save and use for further analysis. The following seven composite variables were derived from the CFA: (a) Autism Characteristics and Severity (ACS), (b) Communication (Comm), (c) Sensory and Repetitive Behaviors, parent report (SRB-P), (d) Sensory and Repetitive Behaviors, teacher report (SRB-T), (e) Reciprocal Social Interaction, parent report (RSI-P), (f) Reciprocal Social Interaction, teacher report (RSI-T), and (g) Fine Motor (FM). The SRB and RSI composites were originally conceptualized as single factors, but examination of the bivariate correlations between indicators provided evidence that the teacher- and parent-rated items were not highly correlated with one another. Therefore, those two constructs were divided into teacher- and parent-rated versions. Through the use of full-information maximum likelihood estimation (Finkbeiner 1979), factor scores could be estimated even in the presence of one or more missing indicator variables, which also increases power by avoiding unnecessary listwise deletion. Table 5 provides the definitions and model fit statistics for the seven composite variables.

The CFA models for all of the composites except for the ACS and FM composites adjusted for the repeated measures structure of the data. The ACS and FM models, being based on three and two indicators, respectively, had too few degrees of freedom to enable the adjustment for clustering. However, the failure to adjust did not impact the factor scores generated by the model for later analysis. Table 6 provides descriptive statistics for the composite variables by group and time point. Oneway ANOVAs

Table 5 Composite variable CFA results

Composite	*n*	χ2 (*df*)	*p*	CFI	RMSEA	SRMR
Autism characteristics and severity (ACS)[a]	477	–	–	–	–	–
Communication (Comm)[b]	477	55.900 (7)	<.001	0.979	0.121	0.018
Sensory and repetitive behaviors, parent-rated (SRB-P)[c]	418	23.778 (2)	<.001	0.968	0.161	0.032
Sensory and repetitive behaviors, teacher-rated (SRB-T)[d]	380	3.465 (2)	.177	0.994	0.044	0.020
Reciprocal social interaction, teacher-rated (RSI-T)[e]	381	3.817 (2)	.201	0.998	0.049	0.010
Reciprocal social interaction, parent-rated (RSI-P)[f]	443	11.973 (5)	.035	0.994	0.056	0.018
Fine motor (FM)[g]	472	6.931 (1)	.020	0.935	0.112	0.131

Model for ACS is just-identified. Model for COMM includes free covariances between PLS and Vineland indicators. Models for ACS and FM do not model repeated measures due to too few degrees of freedom. Model for FM was fit to standardized versions of the indicator variables and fixed both factor loadings to one and the variance of the factor to one

[a] ADOS calibrated severity score, CARS: category rating score total score, SRS-P total score T score, teacher reported

[b] PLS4: expressive communication standard score; Mullen: expressive language standard score; Vineland II: expressive subscale raw. PLS4: auditory comprehension standard score; Mullen: receptive language standard score; Vineland II: receptive subscale raw

[c] RBS-R: IV. Ritualistic behavior: total subscale score, parent reported; RBS-R: I. Stereotyped behavior: total subscale score, parent reported; RBS-R: III. Compulsive behavior: total subscale score, parent reported; RBS-R: V. Sameness behavior: total subscale score, parent reported

[d] RBS-R: I. Stereotyped behavior: total subscale score, teacher reported; RBS-R: V. Sameness behavior: total subscale score, teacher reported; RBS-R: III. Compulsive behavior: total subscale score, teacher reported; RBS-R: IV. Ritualistic behavior: total subscale score, teacher reported

[e] SRS-P social awareness T score, teacher reported; SRS-P social cognition T score, teacher reported; SRS-P social communication T score, teacher reported; SRS-P social motivation T score, teacher reported

[f] SRS-P social awareness T score, parent reported; SRS-P social cognition T score, parent reported; SRS-P social communication T score, parent reported; SRS-P social motivation T score, parent reported; PICS: total score, average of all items

[g] Mullen: scale 3 fine motor standard score calculated, Vineland II: fine motor subscale raw, calculated

Table 6 Descriptives for composite variables by time point and model

	NMS			LEAP			TEACCH		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Time = 1									
ACS**	59	−0.11	0.76	54	0.066	0.765	85	0.381	0.859
COMM***	59	0.214	0.858	54	0.081	1.045	85	−0.403	0.784
SRB (T)*	58	−0.069	0.809	54	−0.176	0.768	85	0.179	0.92
SRB (P)	51	0.025	0.879	49	−0.017	1.03	74	0.169	1.06
RSI (T)	58	0.014	0.999	54	0.24	0.877	85	0.18	0.874
RSI (P)*	58	0.005	0.834	51	−0.056	1.015	80	0.325	0.785
FM**	59	0.01	0.632	54	−0.165	0.812	84	−0.364	0.648
Time = 2									
ACS	54	−0.299	0.928	52	−0.144	0.837	79	0.124	0.866
COMM	54	0.441	0.937	51	0.238	1.102	78	−0.317	0.878
SRB (T)	54	−0.158	0.91	51	−0.212	0.878	78	0.226	1.041
SRB (P)	51	−0.009	0.932	47	−0.141	0.773	59	0.075	0.953
RSI (T)	54	−0.28	1.149	52	−0.152	1.039	78	−0.077	0.926
RSI (P)	52	−0.257	0.969	48	−0.117	1.012	62	0.17	0.845
FM	54	0.44	0.763	51	0.072	0.821	78	−0.183	0.682

* $p < .05$; ** $p < .01$; *** $p < .001$. *ACS* autism characteristics and severity, *COMM* communication, *SRB* sensory and repetitive behaviors parent/teacher rated, *RSI* reciprocal social interaction parent/teacher rated, *FM* fine motor

tested for baseline differences on the composites. The ACS ($p = .0013$), COMM ($p < .001$), SRB-T ($p = .0417$), RSI-P ($p = .0241$), and FM ($p = .0066$) composites did exhibit statistically significant baseline differences. Covariates were added to the model to address baseline differences on composite variables (described below).

Data Analysis

The primary data analytic approach involved the estimation of gain score models in which the outcome was computed as the outcome measured at posttest (Time 2) (based on the child's derived composite variable score) minus the outcome measured at pretest (Time 1) (again based on the derived composite variable score). The gain score model was fit using multilevel or mixed linear models, which adjusted for the clustering of students within classrooms. The models were fit using PROC MIXED in SAS version 9.2. The multilevel model failed to converge in the model for parent-rated reciprocal social interaction, most likely due to a lack of variability at the classroom level. This model was fit using ordinary least squares regression. The advantages of the gain score model were as follows: first, they provide more powerful hypothesis tests than some alternative models due to the purging of between-subjects variance from the outcome; second, they provide some protection against misspecification bias by relieving the analyst of the necessity of correctly specifying the functional form of change over time (e.g., linear vs. non-linear relationships) and the main effects of covariates; and third, they are often simpler to interpret than a repeated measures model. This is because the model only considers change over time. Further in gain score models, pretreatment differences on outcomes are removed via the computation of the gain score; therefore the gain score model completely separates baseline differences from changes over time. Of the three major alternative analytic models for pretest–posttest designs (ANCOVA, repeated-measures models, gain score models), only the latter two produce unbiased estimates when there are baseline differences. The repeated-measures model is capable of estimating relationships between covariates and the baseline outcome as well as change over time, but the tradeoff is increased model complexity. Since none of our research questions involved the baseline time point, the gain score model was a compelling alternative to the repeated-measures approach. We did rerun the analysis using the repeated-measures approach, which required three-level hierarchical linear models, as a sensitivity check to make sure that our results were not overly dependent on the gain-score model specification. The results of the repeated-measures models did not vary from the results of the gain-score models.

A second analysis examined within-model moderators of treatment effect. These models sought to determine the most efficacious treatment model for children with differing covariate patterns. In these models, key covariates were interacted with treatment status and entered into the model, thus allowing estimated treatment effects to vary by the covariate-to-treatment model match. In both cases, missing data were addressed prior to analysis via the application of multiple imputations (MI; Little and Rubin 2002). The imputation model included all the composite variable outcomes and main effects for all predictor variables. In the gain score models, the outcomes included in the imputation model were the gain scores from T1 to T2. Some of the predictor variables were dummy-coded categorical variables, such as race, gender, and education status.

Main Effect Model Results for Pretest to Posttest Change

Results for the main effect models are presented in Table 7. The models were fit to the seven composite variable outcomes described earlier. To assist in addressing pre-treatment differences on composite variables, covariates for the models were parent education, teacher education, gender, and child race (all dummy coded), as well as years of teaching experience, parent stress index total score, Mullen standard score, ADOS severity score, PLS4 score, Leiter BRIEF IQ score, and duration of the school day (half-day vs. full-day program). It must be reiterated that the use of gain scores themselves also allow us to address baseline group differences because they adjust for children's initial status by differencing their pretest and posttest scores, which results in the construction of a gain score. All continuous variables were measured at pretest and were grand-mean centered prior to analysis. Therefore, the model intercepts represent the average change from T1 (pretest) to T2 (posttest) for a "reference" child (i.e., all covariates set to zero) in a NMS classroom. The key variables of interest were dummy variables representing the TEACCH and LEAP treatment models. The NMS model served as the reference group. Post-hoc analyses compared TEACCH to LEAP, TEACCH to NMS, and LEAP to NMS, after adjusting for covariates. Because the latent variable outcomes were standardized to have standard deviations of one, the intercept coefficient as well as the estimates for the pairwise comparisons between models may be interpreted as Cohen's *d*-like effect sizes. For example, an intercept of 0.20 would indicate that the latent outcome increased by approximately 1/5 of a standard deviation from the Time 1 assessment to the Time 2 assessment in NMS (reference) classrooms. A TEACCH versus NMS estimate of 0.5 would indicate that students in TEACCH classrooms experienced an additional average increase of one half of a standard deviation on the latent outcome than the change experienced by students in NMS classrooms.

Primarily, a significant within group effect was found across models for the average change over time variable (see Table 7). The TEACCH group demonstrated significant change over time for 5 of 7 composite variables with the exceptions of teacher- and parent-report of sensory and repetitive behavior (SRB-T and SRB-P, respectively). The LEAP group demonstrated significant change across time on 4 of 7 composites with the exceptions being SRB-T, SRB-P and parent report of child reciprocal social interaction (RSI-P). Finally, the NMS group showed significant change for the following three composite outcome variables: autism severity (ACS), communication (COMM), and fine motor (FM). No statistically significant differences were found between models. The effects of

Table 7 Gain score model results (Time 1 to Time 2)

Variable	ACS	Comm	SRB (T)	SRB (P)	RSI (T)
Avg change for LEAP	−0.378 (0.133)**	0.225 (0.094)*	0.080 (0.260)	0.037 (0.214)	−0.491 (0.196)*
Avg change for NMS	−0.262 (0.117)*	0.236 (0.082)**	0.000 (0.228)	0.180 (0.185)	−0.317 (0.172)
Avg change for TEACCH	−0.375 (0.066)***	0.209 (0.046)***	0.069 (0.139)	−0.105 (0.117)	−0.409 (0.102)***
LEAP versus NMS	−0.116 (0.104)	−0.011 (0.073)	0.080 (0.202)	−0.143 (0.166)	−0.174 (0.152)
TEACCH versus LEAP	0.003 (0.137)	−0.017 (0.096)	−0.011 (0.269)	−0.142 (0.225)	0.082 (0.202)
TEACCH versus NMS	−0.113 (0.127)	−0.027 (0.089)	0.069 (0.245)	−0.285 (0.204)	−0.091 (0.186)

Variable	RSI (P)	FM
Estimate Avg change for LEAP	−0.071 (0.165)	0.279 (0.140)*
Estimate Avg change for NMS	−0.221 (0.144)	0.419 (0.123)***
Estimate Avg change for TEACCH	−0.176 (0.085)*	0.254 (0.070)***
Estimate LEAP versus NMS	0.150 (0.127)	−0.139 (0.109)
Estimate TEACCH versus LEAP	−0.105 (0.172)	−0.025 (0.144)
Estimate TEACCH versus NMS	0.045 (0.158)	−0.164 (0.133)

* $p < .05$; ** $p < .01$; *** $p < .001$. *ACS* autism characteristics and severity, *COMM* communication, *SRB* sensory and repetitive behaviors parent/teacher rated, *RSI* reciprocal social interaction parent/teacher rated, *FM* fine motor. All models excepting RSI (P) account for clustering of students within classrooms. Model for RSI (P) is OLS regression due to multilevel model convergence failure

the covariates were not interpreted because they were only included to reduce confounding bias and increase power.

Moderation Model Results

Results for the moderation models are presented in Table 8. The models included the same covariates from the main effect models, except five covariates were specified as within-treatment model moderators. These covariates included gender, Mullen pretest score, ADOS severity score at pretest, PLS4 score at pretest and parent stress index score measured at pretest. Three significant moderators were discovered. First, a statistically significant Mullen by TEACCH interaction was identified. For the ACS composite, the positive coefficient for this parameter indicated that children with *higher* Mullen scores at pretest made *smaller* gains (i.e., reduction in autism severity) from T1 to T2. As the Mullen is normed to have a standard deviation of 15, a child with a pretest Mullen score one standard deviation higher than the sample mean (a pretest Mullen score of approximately 79.104) would be expected to experience a reduction of autism severity from T1 to T2 by −0.100 standard deviations if placed in a TEACCH classroom (calculated by adding the intercept of −0.208 plus TEACCH main effect of −0.147, plus 15 times the Mullen main effect of −0.010, plus 15 times the Mullen × TEACCH interaction of 0.027). Conversely, a child with a pretest Mullen score of one standard deviation below the sample mean (a pretest Mullen score of approximately 49.10) would be expected to experience a comparatively large reduction in autism severity of −0.610 standard deviations.

A PLS pretest by TEACCH interaction was identified as well. The coefficient of −0.019 indicates that students with high PLS scores would be expected to experience a larger reduction of autism symptoms (ACS composite) from T1 to T2 than would be expected for students with low PLS scores. A PLS by TEACCH interaction was also discovered for the Communication composite, whose positive coefficient of 0.014 indicates that children with high PLS scores in TEACCH classrooms would be expected to exhibit larger gains in communication than would be expected for children with low PLS scores. Finally, a female by LEAP interaction was identified for the Communication outcome. Female students in LEAP classrooms exhibited a smaller gain in communication of 0.057 from T1 to T2, whereas males in LEAP classrooms had a much larger gain of 0.233 standard deviations on the communications composite. However, this interaction effect is difficult to interpret because of the relatively low number of females from LEAP classrooms enrolled in the study (12 of 54 children).

Discussion

This is the first study of which we are aware with the expressed purpose of comparing school-based CTMs for children with ASD. We found that children made gains or reductions in autism characteristics across time irrespective of programmatic type. We did not find change across time

Table 8 Moderation analysis (Time 1 to Time 2) model results

Variable	Tx model	ACS	Comm	SRB (T)	RSI (T)
Intercept (NMS reference group)		−0.208 (0.116)	0.233 (0.087)**	0.028 (0.240)	−0.265 (0.182)
Tx model	N	.	.	.	.
Tx model	L	−0.173 (0.103)	−0.012 (0.077)	0.064 (0.214)	−0.189 (0.161)
Tx model	T	−0.147 (0.124)	0.005 (0.093)	0.098 (0.257)	−0.111 (0.194)
Parent has college degree or higher		−0.046 (0.075)	−0.128 (0.054)*	−0.176 (0.107)	−0.131 (0.099)
Teacher has graduate degree of higher		−0.018 (0.084)	−0.100 (0.064)	−0.183 (0.177)	−0.164 (0.133)
Female		0.081 (0.226)	0.269 (0.160)	0.303 (0.314)	−0.060 (0.291)
Hispanic		−0.168 (0.088)	−0.056 (0.064)	0.043 (0.136)	−0.041 (0.121)
Black		−0.044 (0.126)	−0.103 (0.091)	−0.068 (0.184)	−0.077 (0.169)
Mullen standard score (pretest)		−0.010 (0.008)	0.013 (0.006)*	−0.012 (0.012)	−0.033 (0.011)**
ADOS severity score (pretest)		−0.038 (0.044)	−0.031 (0.031)	−0.048 (0.063)	−0.027 (0.058)
PLS4 standard score (pretest)		0.002 (0.007)	−0.011 (0.005)*	0.004 (0.011)	0.018 (0.010)
BRIEF IQ score (pretest)		−0.004 (0.002)	0.005 (0.002)**	0.001 (0.003)	0.001 (0.003)
Parent stress index total raw score (pretest)		0.004 (0.002)*	−0.001 (0.001)	−0.001 (0.003)	0.005 (0.002)*
Teacher years of experience		−0.012 (0.006)	−0.001 (0.005)	0.005 (0.014)	−0.004 (0.010)
Half day program		0.134 (0.116)	0.007 (0.087)	−0.152 (0.239)	0.037 (0.181)
Dose		0.026 (0.077)	0.001 (0.057)	−0.165 (0.110)	0.033 (0.102)
Female × tx model	N	.	.	.	.
Female × tx model	L	−0.027 (0.280)	−0.443 (0.201)*	−0.285 (0.405)	−0.111 (0.368)
Female × tx model	T	−0.191 (0.274)	−0.161 (0.194)	−0.219 (0.389)	−0.414 (0.358)
Mullen (T1) × tx model	N	.	.	.	.
Mullen (T1) × tx model	L	0.004 (0.012)	−0.005 (0.009)	0.021 (0.018)	0.020 (0.016)
Mullen (T1) × tx model	T	0.027 (0.011)**	−0.010 (0.008)	0.005 (0.016)	0.024 (0.014)
ADOS Severity (T1) × tx model	N	.	.	.	.
ADOS Severity (T1) × tx model	L	0.039 (0.060)	0.053 (0.043)	0.110 (0.087)	0.106 (0.079)
ADOS Severity (T1) × tx model	T	−0.039 (0.060)	0.045 (0.042)	−0.022 (0.083)	−0.015 (0.079)
PLS4 (T1) × tx model	N	.	.	.	.
PLS4 (T1) × tx model	L	0.004 (0.010)	0.003 (0.007)	−0.012 (0.015)	−0.011 (0.013)
PLS4 (T1) × tx model	T	−0.019 (0.009)*	0.014 (0.007)*	0.002 (0.014)	−0.016 (0.013)

* $p < .05$; ** $p < .01$; *** $p < .001$. L = LEAP, T = TEACCH, N = NMS. *ACS* autism characteristics and severity, *COMM* communication, *SRB* sensory and repetitive behaviors (teacher rated), *RSI* reciprocal social interaction (teacher rated). All models account for clustering of students within classrooms

for the parent- or teacher-reported sensory and repetitive behavior composite for any of the models; however, this may reflect that these behaviors are by their nature resistant to change and there are few evidence-based interventions specifically designed to target these characteristics of autism (Bodfish 2004; Boyd et al. 2012). It may be somewhat surprising that significant change across time was not found in LEAP for parent report of social interaction given that the primary method of instruction is peer-mediated instructional strategies. This finding may reflect the higher expectations of these parents for their children's social skills as they observe on a day-to-day basis how the social performance of their children compare to typically developing peers; yet, we did find change for teacher report of children's social skills.

In addition, we found that children's pretest Mullen and PLS scores moderated the effects of TEACCH on children's autism severity, with children with lower Mullen but higher PLS scores at pretest having better outcomes on this composite. Higher PLS scores also moderated the effects of TEACCH on children's communication outcomes. It is consistent with other treatment research studies that children with higher baseline scores tend to have better outcomes (Sallows and Graupner 2005). However, the Mullen finding is of interest because it suggests that children enrolled in TEACCH classrooms with lower versus higher cognitive ability showed more improvement in autism severity. This finding could be attributable to children with lower cognitive abilities likely having more severe symptoms of autism and thus more room for improvement; or it may suggest that some of the environmental and behavioral supports used in TEACCH are more beneficial to children with greater cognitive impairments.

Gender appeared to moderate the communication outcomes of children in LEAP classrooms, with girls showing less improvement for this composite variable, yet, interpretation of this outcome must be cautioned because of the relatively low numbers of females in the study. Still the overall findings, which demonstrate change across time and no model differences, may reflect the importance of general programmatic quality in promoting the positive development of children with ASD. The early childhood literature is replete with studies demonstrating that classroom quality is an important predictor of typically developing children's social, language and academic outcomes (Burchinal et al. 2000; Mashburn et al. 2008; Pianta et al. 2002; Rimm-Kaufman et al. 2005). This concept has not often been measured or considered in autism intervention research, possibly because there have been fewer studies focused on school-based CTMs. It stands to reason that teachers may be better able to implement evidence-based practices, including specific CTMs, when the classroom has a certain level of foundational quality. Our quality measure included such domains as classroom organization, positive instructional climate, collaborative interdisciplinary teaming, and family involvement, and these as well as other areas may be essential building blocks upon which other practices can then be successfully layered. However, we cannot draw firm conclusions without a "lower" quality comparison condition.

It is also important to restate that studies have found positive effects for school-based implementation of TEACCH (Tsang et al. 2007) and LEAP (Strain and Bovey 2011). In particular, the Strain and Bovey study was a large-scale, well-designed evaluation study of LEAP that demonstrated its superiority to a control condition. The difference between their study and the current one is that their comparison condition consisted of teachers having access to the LEAP treatment manual and training materials (i.e., a low-fidelity LEAP), whereas our comparison conditions were comprised of another active treatment (TEACCH) and high quality control classrooms (NMS). In treatment efficacy research, the control condition plays a large role in detecting between group differences (Mohr et al. 2009). Thus, it is possible that our NMS "control" group really acted as another active treatment because all the classrooms were high quality. The study may not have been sufficiently powered to detect differences between three active treatment conditions. The dissimilar findings between the current study and prior studies of TEACCH and LEAP also may reflect the difficulty of sustaining intervention effects when the model developer is not directly involved in the research. For instance, Reichow and Wolery (2009) found that the effects of the Lovaas Young Autism Project on child IQ was moderated by training personnel, with larger changes in IQ found if the study personnel had been trained in the UCLA model, which is the origin of the Lovaas approach. While not involving the model developer may help to reduce researcher bias, it could also be that there are intangible aspects of the intervention that are more easily conveyed when the purveyor is directly involved. However, our fidelity measures indicated that TEACCH and LEAP teachers strongly adhered to the components of the model that could be observed, and did so consistently across the school year (Coman et al. 2013). This concept of model developers not being involved in the research also moves the current study into the area of effectiveness versus efficacy research.

Within the RCTs literature, the term "pragmatic randomized trial" is used to describe studies that combine elements of both efficacy and effectiveness trials (Marchand et al. 2011; Zwarenstein and Treweek 2009). This study included elements of efficacy trials in that it purposefully enrolled high quality classrooms to minimize pretest differences, and provided booster training to LEAP and TEACCH teachers to increase and/or maintain their fidelity of implementation. Having to combine elements of both study designs, because LEAP and TEACCH are already widely used in practice, may have affected study outcomes. However, pragmatic randomized trials also contribute to generalizability because they often involve community-based intervention agents. The study findings also could reflect limitations with the use of the fidelity measures to screen in/select classrooms (e.g., they may not have fully captured the active ingredients of the model that best reflect adherence), or the outcome measures used may not have been the most sensitive to detect treatment differences. Finally, the study findings may simply reflect that high quality teachers are aware of and use similar practices to educate children with autism. In fact, this crossover of classroom practices was made evident through our fidelity/quality data as well as CPI data. While a likely limitation, we believe this overlap reflects the real-world heterogeneity found when studying classroom-based practices; i.e., school-based practitioners are trying out and using a number of strategies to educate children with ASD. It should be pointed out that although the CPI data indicated that teachers self-reported the use of a variety of supplemental practices (e.g., pivotal response training), we did not collect observational data to confirm those self-reports or measure the quality of implementation of those ancillary practices.

There are also obvious limitations with the use of quasi-experimental designs, which could account for study outcomes. First, we cannot rule out that the findings are the result of sheer developmental maturation or selection bias issues. For example, we had to use raw versus standard scores for some measures (e.g., CARS), which may make these measures more susceptible to the effects of developmental maturation. However, the difficulty of randomization when studying existing school-based CTMs necessitated the use of a quasi-experimental study, and we put appropriate protocols in place, such as screening classrooms for quality and monitoring fidelity of implementation across time, to ensure a rigorous study design. Further, selection bias is likely an issue with most school-based studies because one is often dependent upon school officials to nominate classrooms/teachers to participate in the study, and it could be that officials self-select higher quality programs. Second, there were pre-treatment differences between groups; however, the use of a gain score analysis approach allowed us to address these initial differences. Third, assessors were not blind to children's group assignment. It would have been difficult to maintain assessor blindness because the majority of assessments were conducted in children's schools and assessors could see the stark contrasts in the physical layout as well as types of children in TEACCH versus LEAP classrooms (all children with ASD versus more typically developing children). We attempted to counter this bias by not analyzing any child outcome data until all data collection had been completed. Finally, the NMS classrooms in our study may not reflect the real-world heterogeneity found in actual practice. These NMS classrooms may represent the "best" of standard practice and may differ substantially from the modal level of quality that reflects "business-as-usual" classroom practices.

Implications and Future Research

All three programs were found to produce statistically significant changes in children's outcomes across the school year. This finding may shift the field's thinking around CTMs designed for students with ASD. Perhaps it is not the *unique* features of the models that most contribute to child gains; instead it is the *common* features of the models that most influence child growth. In other fields, such as clinical and counseling psychology, the concept of "common factors" is well used and understood when describing therapeutic interventions for clients. These are intervention components that should apply to all clients and should be in place for effective practice to occur (e.g., trusting relationship, Deegear and Lawson 2003; Luborsky 1995). The common factor theory may well apply to interventions for children with ASD (Odom et al. 2012). Early analysis of the overlap of scores on the fidelity measures indicate that perhaps those components common to the intervention approaches (e.g. classroom organization, teacher interaction with students and families) account for outcomes more than components that are unique to each approach (e.g. peer-mediated instruction in LEAP classrooms, structured work systems in TEACCH classrooms). A more complex analysis of the fidelity and quality measures and outcomes will allow for further exploration of this common factor theory and its application to CTMs for young children with ASD.

Examining the treatment effects of CTMs on child level outcome variables is important, however when selecting educational models for implementation, caregivers, teachers, and school administrators must evaluate multiple features of the model and its impact on multiple variables. These include outcomes for caregivers whose children (e.g. mental health, stress) are enrolled in specific school-based programs, their impact on teachers (e.g. teacher burnout), and the financial costs to the district to adopt and implement a specific model. For instance, by design, LEAP classrooms are operated as half-day programs and if children are able to receive similar outcomes in half-day as full-day programs, then the cost-benefits may factor into school administrators' decision to select a particular CTM. Because there were no differential effects for child outcomes, further exploration of the effect or impact of these classroom programs on these and other variables is an important next step. With these supplemental analyses, in addition to further exploration of treatment moderators, stakeholders will be better informed when selecting and implementing CTMs in public school settings.

Acknowledgments The research reported here was supported by the Institute of Education Sciences, US Department of Education through Grant R324B070219 awarded to UNC-Chapel Hill. The opinions expressed represent those of the authors and do not represent views of the Institute or the US Department of Education. At the time the study was conducted, Dr. Matthew McBee was at the UNC-Chapel Hill, Dr. Anibal Gutierrez was at the University of Miami and Dr. Laurie Sperry was at the University of Colorado-Denver. Finally, we wish to thank all of the families, teachers and other school personnel who participated in this study.

References

Bodfish, J. W. (2004). Treating the core features of autism: Are we there yet? *Mental Retardation and Developmental Disabilities Research Reviews, 10*, 318–326.

Bodfish, J. W., Symons, F. W., & Lewis, M. H. (1999). *The repetitive behavior scale*. Western Carolina Center Research Reports.

Boyd, B. A., McDonough, S., & Bodfish, J. W. (2012). Evidence-based behavioral interventions for repetitive behaviors in autism. *Journal of Autism and Developmental Disorders, 42*, 1236–1248.

Boyd, B. A., Odom, S., Humphreys, B., & Sam, A. (2010). Infants and toddlers with ASD: Early identification and early intervention. *Journal of Early Intervention, 32*, 75–98.

Burchinal, M., Peisner-Feinberg, E., Bryant, D., & Clifford, R. (2000). Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Journal of Applied Developmental Sciences, 4*, 149–165.

Classroom Practice Indicator. (2007). Unpublished instrument. Project team of "*Comparison of two comprehensive treatment models for preschool-aged children with autism and their families*". Chapel Hill, NC: UNC-CH.

Coman, D. C., Alessandri, M., Gutierrez, A., Novotny, S., Boyd, B. A., Hume, K., et al. (2013). Commitment to classroom model philosophy and burnout symptoms among high fidelity teachers implementing preschool programs for children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*, 345–360.

Constantino, J. N. (2002). *The social responsiveness scale*. Los Angeles: Western Psychological Services.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model. *Pediatrics, 125*, 17–23.

Deegear, J., & Lawson, D. M. (2003). The utility of empirically supported treatments. *Professional Psychology: Research and Practice, 34*, 271–277.

Delgado, C., Mundy, P., & Block, J. (2001). *Pictorial infant communication scale (PICS) version 1.3*. Coral Gables: University of Miami.

Dingfelder, H., & Mandell, D. (2011). Bridging the research to practice gap in autism intervention: An application of diffusion of innovation theory. *Journal of Autism and Developmental Disorders, 41*, 597–609.

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral intervention at school for 4- to 7-year-old children with autism. *Behavior Modification, 26*, 49–68.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology, 38*(3), 439–450.

Farmer, C., Lecavalier, L., Yu, S., Arnold, L. E., McDougle, C., et al. (2012). Predictors and moderators of parent training efficacy in a sample of children with autism spectrum disorders and serious behavioral problems. *Journal of Autism and Developmental Disorders, 42*, 1037–1044.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika, 44*, 409–420.

Hume, K., Boyd, B., McBee, M., Coman, D., Gutierrez, A., Shaw, E., et al. (2011). Assessing implementation of comprehensive treatment models for young children with ASD: Reliability and validity of two measures. *Research in Autism Spectrum Disorders, 11*, 1430–1440.

Hume, K., & Odom, S. (2007). Effects of an individual work system on the independent functioning of students with autism. *Journal of Autism and Developmental Disorders, 37*, 1166–1180.

Hume, K., Plavnick, J., & Odom, S. (2012). Promoting task accuracy and independence in students with autism across educational setting through the use of individual work systems. *Journal of Autism and Developmental Disorders, 42*, 2084–2099.

Jennett, H., Harris, S., & Mesibov, G. (2003). Commitment to philosophy, teacher efficacy, and burnout among teachers of children with autism. *Journal of Autism and Developmental Disorders, 33*, 583–593.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, 47*, 611–620.

Landa, R., Holman, K., O'Neill, A., & Stuart, E. (2011). Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 52*, 13–21.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *The autism diagnostic observation schedule (ADOS)*. Los Angeles: Western Psychological Corporation.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9.

Luborsky, L. (1995). Are common factors across different psychotherapies the main explanation for the Dodo bird verdict that "everybody has won so we shall claim prizes?". *Clinical Psychology: Science and Practices, 2*, 106–109.

Marchand, E., Stice, E., Rohde, P., & Becker, C. (2011). Moving from efficacy to effectiveness trials in prevention research. *Behaviour Research and Therapy, 49*, 32–41.

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732–749.

Mesibov, G. B., & Shea, V. (2010). The TEACCH program in the area of evidence-based practice. *Journal of Autism and Developmental Disorders, 40*, 570–579.

Mesibov, G. B., Shea, V., & Schopler, E. (2005). *The TEACCH approach to autism spectrum disorders*. New York: Kluwer Academic/Plenum.

Mohr, D., Spring, B., Freedland, K., Beckner, V., Arean, P., Hollon, S., et al. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics, 78*, 275–284.

Mullen, E. M. (1995). *Mullen scales of early learning* (AGS ed.). Los Angeles: Western Psychological.

Odom, S. L., Boyd, B., Hall, L. J., & Hume, K. (2010a). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*, 425–436.

Odom, S. L., Brown, W. H., Frey, T., Karasu, N., Smith-Carter, L., & Strain, P. (2003). Evidence-based practices for young children with autism: Evidence from single-subject research design. *Focus on Autism and Other Developmental Disabilities, 18*, 176–181.

Odom, S. L., Collet-Klingenberg, L., Rogers, S., & Hatton, D. (2010b). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure, 54*, 275–282.

Odom, S. L., Hume, K., Boyd, B. A., & Stabel, A. (2012). Moving beyond the intensive behavioral treatment versus eclectic dichotomy: Evidence-based and individualized programs for learners with ASD. *Behavior Modification, 36*, 270–297.

Ozonoff, S., & Cathcart, K. (1998). Effectiveness of a home program intervention for young children with autism. *Journal of Autism and Developmental Disorders, 28*, 25–32.

Panerai, S., Ferrante, L., Caputo, V., & Impellizzeri, C. (1998). Use of structured teaching for treatment of children with autism and severe and profound mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 33*, 367–374.

Panerai, S., Ferrante, L., & Zingale, M. (2002). Benefits of the treatment and education of autistic and communication handicapped children (TEACCH) programme as compared with a non-specific approach. *Journal of Intellectual Disability Research, 46*, 318–327.

Panerai, S., Zingale, M., Trubia, G., Finocchiaro, M., Zuccarello, R., Ferri, R., et al. (2009). Special education versus inclusive education: The role of the TEACCH program. *Journal of Autism and Developmental Disorders, 39*(6), 874–882.

Pianta, R. C., LaParo, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *Elementary School Journal, 102*, 225–238.

Professional Development in Autism Center. (2008). *PDA program assessment*. Seattle, WA: PDA Center at University of Washington.

Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the UCLA young autism project model. *Journal of Autism and Developmental Disorders, 39*, 23–41.

Rimm-Kaufman, S., LaParo, K., Downer, J., & Pianta, R. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal, 105*, 377–394.

Roid, G., & Miller, L. (1997). *Leiter international performance scale—revised: Examiners manual*. Wooddale, IL: Stoelting Co.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire (SCQ)*. Los Angeles, CA: Western Psychological Services.

Sallows, G., & Graupner, T. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*, 417–438.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The childhood autism rating scale*. Los Angeles, CA: Western Psychological Services.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation, 105*, 269–285.

Sox, H., & Greenfield, S. (2009). Comparative effectiveness research: A report from the institute of medicine. *Annals of Internal Medicine, 15*, 203–205.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.

Stahmer, A. C., & Ingersoll, B. (2004). Inclusive programming for toddlers with autism spectrum disorders: Outcomes form the children's toddler school. *Journal of Positive Behavior Interventions, 6*, 67–82.

Strain, P., & Bovey, E. (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.

Strain, P. S., & Hoyson, M. (2000). The need for longitudinal, intensive social skills intervention: LEAP follow-up outcomes for children with autism. *Topics in Early Childhood Special Education, 20*, 116–122.

Strain, P. S., Hoyson, M., & Jamieson, B. (1985). Normally developing preschoolers as intervention agents for autistic-like children: Effects on class deportment and social interaction. *Journal of the Division for Early Childhood, 9*, 105–109.

Tsang, S., Shek, D., Lam, L., Tang, F., & Cheung, P. (2007). Brief report: Application of the TEACCH program on Chinese preschool children with autism. Does culture make a difference? *Journal of Autism and Developmental Disorders, 37*, 390–396.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scales* (4th ed.). San Antonio, TX: Psychological Corporation.

Zwarenstein, M., & Treweek, S. (2009). What kind of randomized trials do we need? *Canadian Medical Association Journal, 180*, 999–1000.