Research in Developmental Disabilities 35 (2014) 3326–3344

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Comparison of behavior analytic and eclectic early interventions for young children with autism after three years

Jane S. Howard [a,b,1,*], Harold Stanislaw [a], Gina Green [c], Coleen R. Sparkman [b,1], Howard G. Cohen [d]

[a] California State University, Stanislaus, Psychology Department, 1 University Circle, Turlock, CA 95382, USA
[b] The Kendall Centers/Therapeutic Pathways, Modesto, CA 95354, USA
[c] Association of Professional Behavior Analysts, 6977 Navajo Road #176, San Diego, CA 92119, USA
[d] Valley Mountain Regional Center, 702 North Aurora St, Stockton, CA 95202, USA

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form 6 August 2014
Accepted 12 August 2014
Available online

Keywords:
Autism
Early intervention
Applied behavior analysis
Eclectic treatment
Outcomes
Longitudinal studies

ABSTRACT

In a previous study, we compared the effects of just over one year of intensive behavior analytic intervention (IBT) provided to 29 young children diagnosed with autism with two eclectic (i.e., mixed-method) interventions (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005). One eclectic intervention (autism programming; AP) was designed specifically for children with autism and was intensive in that it was delivered for an average of 25–30 h per week ($n = 16$). The other eclectic intervention (generic programming; GP) was delivered to 16 children with a variety of diagnoses and needs for an average of 15–17 h per week. This paper reports outcomes for children in all three groups after two additional years of intervention. With few exceptions, the benefits of IBT documented in our first study were sustained throughout Years 2 and 3. At their final assessment, children who received IBT were more than twice as likely to score in the normal range on measures of cognitive, language, and adaptive functioning than were children who received either form of eclectic intervention. Significantly more children in the IBT group than in the other two groups had IQ, language, and adaptive behavior test scores that increased by at least one standard deviation from intake to final assessment. Although the largest improvements for children in the IBT group generally occurred during Year 1, many children in that group whose scores were below the normal range after the first year of intervention attained scores in the normal range of functioning with one or two years of additional intervention. In contrast, children in the two eclectic treatment groups were unlikely to attain scores in the normal range after the first year of intervention, and many of those who had scores in the normal range in the first year fell out of the normal range in subsequent years. There were no consistent differences in outcomes at Years 2 and 3 between the two groups who received eclectic interventions. These results provide further evidence that intensive behavior analytic intervention delivered at an early age is more likely to produce substantial improvements in young children with autism than common eclectic interventions, even when the latter are intensive.

* Corresponding author at: California State University, Stanislaus, Psychology Department, 1 University Circle, Turlock, CA 95382, USA. Tel.: +1 209 667 3386; fax: +1 209 993 8192.

E-mail addresses: jhoward@csustan.edu, janeshoward@mac.com (J.S. Howard), hstanislaw@csustan.edu (H. Stanislaw), ggreen@apbahome.net (G. Green), csparkman@tpathways.org (C.R. Sparkman).

[1] Address: PO Box 5157, Modesto, CA 95352, USA.

http://dx.doi.org/10.1016/j.ridd.2014.08.021

1. Introduction

The past two decades have seen increased interest in early intervention for children diagnosed with autism spectrum disorder (hereafter, "autism") among researchers, policymakers, funding sources, and consumers. Following publication of the Lovaas study in 1987, a number of researchers began evaluating the effects of intensive, comprehensive early intervention using applied behavior analysis (ABA) methods. Various ABA models for treating children with autism have been proposed, but many behavior analytic researchers agree that genuine early intensive ABA treatment programs have certain key features in common: (a) individualized, comprehensive intervention that addresses all skill domains; (b) use of multiple behavior analytic procedures (not just discrete-trial procedures or "naturalistic" techniques) to build new repertoires and reduce behaviors that interfere with skill acquisition and effective functioning; (c) direction and oversight by one or more professionals with advanced training in ABA and experience with young children with autism; (d) reliance on typical developmental sequences to guide selection of treatment goals; (e) parents and other individuals trained by behavior analysts to serve as active co-therapists; (f) intervention that is initially one-to-one, transitioning gradually to a group format as warranted; (g) intervention that often begins in homes or specialized treatment centers but is also delivered in other environments, with gradual, systematic transitions to regular schools when children develop the skills required to learn in those settings; (h) planned, structured intervention provided for a minimum of 20–30 h per week with additional hours of informal intervention provided throughout most other waking hours, year round; (i) intensive intervention beginning in the preschool years and continuing for at least 2 years (Eldevik et al., 2010; Green, Brennan, & Fein, 2002).

Substantial research has documented the effectiveness of treatments that incorporate all of the foregoing features. Eight prospective studies used comparison- or control-group designs to evaluate some variation of the Lovaas/UCLA model of early intensive ABA intervention for children with autism (Cohen, Amerine-Dickens, & Smith, 2006; Eikeseth, Smith, Jahr, & Eldevik, 2002; Eikeseth, Smith, Jahr, & Eldevik, 2007; Eldevik, Hastings, Jahr, & Hughes, 2012; Eldevik, Eikeseth, Jahr, & Smith, 2006; Lovaas, 1987; Sallows & Graupner, 2005; Smith, Groen, & Wynn, 2000). In another three studies, the ABA intervention was designed and overseen by professional behavior analysts not affiliated with Lovaas, and the ABA intervention differed somewhat from the Lovaas model (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Remington et al., 2007; Zachor, Ben-Itzchak, Rabinovitch, & Lahat, 2007). Outcomes from those 11 studies varied and some children had larger improvements than others. In the large majority of cases, however, the mean change scores achieved by children receiving intensive ABA treatment exceeded the mean change scores for similar children in control or comparison groups who received less intensive ABA treatment, intensive or non-intensive treatment using a mixture of methods or therapies ("eclectic" treatment), or "treatment as usual" (i.e., standard early intervention or special education services). Additionally, compared to children who received other types of treatment, children who received early intensive ABA treatment were more likely to achieve post-treatment scores on one or more standardized measures that were in the normal range, and were more often placed in regular classrooms (for reviews and analyses, see Eikeseth, 2009; Eldevik et al., 2009, 2010; Green, 2011; National Autism Center, 2009; Reichow & Wolery, 2009; Rogers & Vismara, 2008).

Despite the evidence from multiple studies and meta-analyses favoring intensive ABA treatment for autism over other models of early intervention, a number of questions persist. One is whether other types of early intervention delivered with comparable intensity and individualization can produce outcomes comparable to ABA. Perhaps the most common alternative early intervention approach involves a mixture of methods drawn from ABA, speech-language pathology, occupational therapy (especially sensory integration techniques), developmental psychology, and autism-specific approaches. That model, which has been characterized as "eclectic" intervention, is widely available in the United States and elsewhere.

At least three studies have compared eclectic and ABA interventions directly. Eikeseth et al. (2002) studied children with autism who entered treatment at ages 4–7 years (M = 5.5 years), slightly older than children in most of the other studies of early intensive behavioral intervention. One group (n = 13) received Lovaas-model ABA treatment for 28 h per week, while a second group (n = 12) received eclectic intervention for 29 h per week. There were no significant differences between the groups when treatment began. Both forms of treatment were delivered in public school classrooms. After 1 year, the ABA treatment group had gained an average of 17 points on IQ test scores, 13 points on tests of language comprehension, 27 points on tests of expressive language, and 11 points on an adaptive behavior scale. The eclectic treatment group had average gains of only 4 points on IQ tests and 1 point on language tests, and no change in adaptive behavior. A follow-up study conducted when those children were 8 years old found that after about 3 years of treatment, the ABA treatment group had gained an average of 25 IQ points and 9–20 points on adaptive behavior scales in comparison to baseline. The eclectic intervention group had a mean gain of only 7 points on IQ tests, and declines of 6–12 points on adaptive behavior assessments (Eikeseth et al., 2007).

A study we published previously involved a comparison of intensive ABA intervention with two different eclectic intervention models (Howard et al., 2005). Twenty-nine preschool children with autism received early intensive behavior analytic intervention (IBT), 16 received intensive eclectic intervention designed for children with autism (designated the autism programming, or AP, group), and an additional 16 received typical non-intensive, eclectic early intervention services (designated the generic programming, or GP, group). All children began intervention prior to 48 months of age and received treatment for an average of 14 months. They were placed in treatment groups on the basis of parental preferences and education team decisions, and evaluated pre-treatment and annually thereafter by professionals who were neither involved in nor employed by any of the treatment programs. The three groups were shown to be similar on key variables when treatment began. After 14 months of intervention, mean scores on standardized tests of intellectual, communication, and adaptive skills were significantly higher for children in the IBT group than for children in the other two groups. Children in the IBT group had an average standard IQ score of 90, compared to 62 and 69 for children in the AP and GP groups, respectively. Developmental trajectories for most measures accelerated markedly over the 14 months of treatment for children in the IBT group, while the trajectories for children in the other two groups remained flat or declined.

For the present study, we followed children who participated in the 2005 study through an additional 2 years of treatment. We focused on four questions: (a) did Year 1 differences in the cognitive, language, and adaptive behavior scores of children in the three groups persist? (b) Did differences in the developmental trajectories of the three groups at Year 1 change during Years 2 and 3? (c) How many children in each group had standardized test scores in the normal range after 2 or 3 years of treatment? (d) To what extent were outcomes at Year 1 correlated with outcomes at Years 2 and 3?

2. Method

2.1. Participants

The same 61 children who participated in the Howard et al. (2005) study participated in this follow-up. Characteristics of the groups at intake are reported in Howard et al. (2005).[2]

Assessments were conducted 1–3 years after treatment began, but not all skill domains were assessed each year with every child. (See Section 3.1 for number of assessments available for each group at intake and at Years 1–3.) In particular, one child in the GP group and one child in the IBT group did not receive any assessments after the first year of treatment. Nonetheless, scores for all 61 children were retained for the present analyses to permit evaluations of outcomes that were not included in our 2005 publication.

2.2. Treatments

Information about the treatments participants received, school placements, and number of hours and services authorized during Years 2 and 3 was obtained through file review.

2.2.1. Intensive behavior analytic treatment (IBT)

This treatment was designed and delivered by personnel in a California non-public agency that provides ABA services to children with autism. Treatment was directed by the first author, a Board Certified Behavior Analyst-Doctoral® (BCBA-D®) and licensed psychologist, and the fourth author, a licensed speech-language pathologist. Programs were supervised by Board Certified Behavior Analysts® (BCBAs®) and other staff with master's degrees in psychology or special education and some training in ABA. They were supported by staff who were either Board Certified Assistant Behavior Analysts® (BCaBAs®) or who had bachelors degrees, most of whom were enrolled in graduate programs in ABA and related areas. Treatments were delivered to children by behavior technicians working under the supervision of the clinical staff. Behavior technicians began delivering treatment only after they had passed competency-based performance evaluations; thereafter, they were directly observed and received written or oral feedback on their implementation of behavior change protocols from their clinical supervisors an average of once or twice each week.

To varying degrees, all parents helped support treatment outside of formal treatment hours. Parent training initially focused on teaching instruction-following, promoting spontaneous language, re-directing nonfunctional repetitive behavior, managing interfering behaviors, and building skills such as toileting, dressing, and independent play. Parents were also trained to implement behavior analytic procedures that were designed to increase success in activities relevant to health and self-care, such as cooperating with medical and dental care procedures and participating in sports and other community activities.

Treatment was delivered in multiple settings, including homes, treatment centers, community settings, and regular preschool and elementary school classrooms. Treatment protocols utilized the full range of behavior analytic procedures, customized to each child's level of functioning, preferences, family circumstances, and treatment goals. Each child received an average of 35–40 h of treatment per week. The adult:child ratio during Year 1 was 1:1, but during Years 2 and 3 the ratio was gradually decreased (e.g., to 1:2 or 1:3, and then to one adult per small group of children), depending on progress and treatment targets. For further details, see Howard et al. (2005).

---

[2] While assembling the data for this study, we uncovered several errors in data reported in our 2005 paper. Most were minor (e.g., 1-month errors in the child's age), but the baseline scores of one child in the GP group were reported incorrectly as Year 1 scores, and the Year 1 scores of another child in the GP group were reported as baseline scores. Correcting those errors had virtually no impact on the conclusions that were drawn in the 2005 paper; all 107 of the statistical tests reported as not significant in 2005 remained non-significant, and 40 of the 43 findings that were reported as statistically significant in 2005 remained so. The three exceptions were for group differences that were only marginally significant in the 2005 publication: the difference at intake between the mean nonverbal age equivalents for the AP and GP groups changed from $p = 0.04$ to $p = 0.07$; the difference at follow-up between the mean motor standard scores of the IBT group and the two comparison groups changed from $p = 0.04$ to $p = 0.06$; and, when the mean self-help skills learning rates before and after treatment were compared, the difference between the IBT group and the two comparison groups changed from $p = 0.05$ to $p = 0.07$. Revised tables reporting all corrections are available as supplementary materials.

Data from standardized assessments as well as direct observation and measurement of target behaviors guided decision-making about the distribution of treatment hours across targets and settings. Initial treatment targets focused on foundational repertoires (e.g., attending, imitating vocal and motor sequences, following spoken directions, receptive and expressive labeling, initiating requests, tolerating change, etc.) that are often absent or at low levels in children with autism. Treatment targets during Years 2 and 3 generally focused on advanced cognitive, social, play, self-care, academic, and communication skills (for example, see Fischer, Howard, Sparkman, & Moore, 2009). More complex interactions involving peers and siblings generally occurred during Years 2 and 3 than in Year 1. On average, children in the IBT group had more than 200 goals on their annual individualized education programs (IEPs).

When children acquired the skills necessary to benefit from small group instruction (e.g., learning through observing the behavior of others, language skills close to the level of instruction, low levels of problem behaviors, independent communication of basic needs), they were placed in preschool or kindergarten programs for typically developing children for up to 15 h per week. Each child was accompanied by a behavior technician who used a variety of behavior analytic approaches, including self-management and behavioral contracting procedures, to arrange opportunities to prompt and reinforce behavior targets in order to promote skill acquisition and generalization across settings. The clinical supervisor directing the intervention also provided training and consultation to parents, teachers, and other professionals. Sample behaviors targeted in the regular classrooms included following instructions from classroom teachers and aides, engaging in classroom routines, and interacting with peers. Time spent with typically developing peers was gradually increased based on skill acquisition, maintenance and generalization of skills, and level of problem behaviors. Most children did not enter kindergarten until age 6.

2.2.2. Autism programming (AP) and generic programming (GP)

Brief descriptions of the AP and GP interventions are presented next; for details see Howard et al. (2005). The AP programs were designed specifically for children with autism. Intervention procedures were drawn from the Training and Education of Autistic and Related Communication Handicapped Children (TEACCH) approach, sensory integration therapy, commercially available programs (e.g., the Picture Exchange Communication System; Bondy & Frost, 1994), and some behavior analytic procedures, such as discrete-trial procedures. Children in this group received an average of 25–30 h of intervention per week in public school classrooms with staffing ratios of 1:1 or 1:2. Thus, the AP programs provided eclectic intervention at an intensity that was comparable to IBT.

The GP intervention was delivered in special education classrooms that served children with a variety of diagnoses and educational needs. Programming that was described as "developmentally appropriate" and "language rich" was provided for an average of 15–17 h per week, with slightly more hours as children approached age 6. Adult:child ratios averaged 1:6.

Approximately one third of the children in both the AP and GP treatment groups received "pull out" speech therapy sessions of less than 30 minutes once or twice a week during Years 1 and 2. About 20% of the children in both groups received some services in general education classrooms, which often included such activities as lunch, physical education, or recess. On average, each child in the AP and GP groups had fewer than 15 goals on his/her annual IEP.

2.2.3. Summary

All of the children in the IBT group and the majority in the two eclectic treatment groups had similar placements and programming during Years 2 and 3 as in Year 1. Some children changed from one eclectic treatment to the other after Year 1, while the Year 2 and/or Year 3 intervention was not available for a few AP and GP children. This information is summarized in Fig. 1, which is similar to a Sankey diagram. Sankey diagrams are used in engineering (see Schmidt, 2008, for an overview) and vary the width of each arrow in proportion to the number represented by that arrow. Thus, in the GP treatment group, the arrow leading from GP treatment in Year 1 to AP treatment in Year 2 (which represents $n = 3$ children) is three times as wide as the arrow leading from AP treatment in Year 2 back to GP treatment in Year 3 (representing $n = 1$ child).

2.3. Design

We utilized a between-groups design to compare performances of children in the IBT group with those of children in the two eclectic treatment groups at intake and at followup assessments about 1–3 years later. As reported in Howard et al. (2005), the three groups of children were substantially similar on most key variables at intake. The only significant differences were in mean chronological ages and parents' education, which were controlled for statistically (see Section 2.3.2 below).

2.3.1. Dependent measures

The principal dependent measures in this study were scores on full-scale IQ tests (cognitive skills), measures of language development, and adaptive behavior scales (composite scores as well as communication, self-help, and social skills scores). Scores on nonverbal IQ tests, receptive and expressive communication skills assessments, and motor skills were also analyzed. Since these latter skills were often not measured in Year 3, we report the Year 2 scores if Year 3 scores were not available.

All intake and follow-up assessments were conducted by experienced, qualified examiners who were not involved in treating any children in any of the groups. Assessments were conducted in the child's home, in the examiner's office, at a

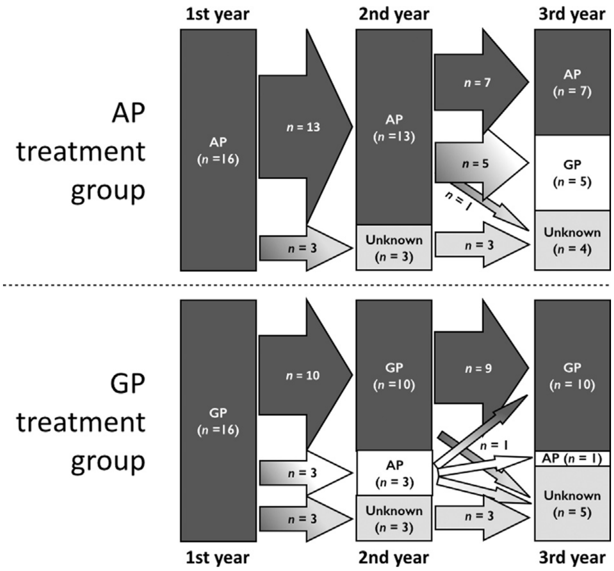

Fig. 1. Movement of children between AP and GP treatments by year. Children in the IBT treatment group had the same treatment all three years. Agency files did not report the type of treatment that was received during Years 2 and 3 for some of the children who initially received the AP or GP treatments.

Table 1
Age (in months) at each assessment, and interval between intake and each subsequent assessment.

Measure	IBT		AP		GP		IBT mean minus AP/GP mean	AP mean minus GP mean
	M	SD	*M*	SD	*M*	SD		
Age at diagnosis	30.07	5.30	39.31	5.52	34.94	5.18	−7.06**	4.38*
Age at intake testing	30.86	5.16	37.44	5.68	34.75	4.80	−5.23**	2.69
Age at Year 1 follow-up	45.24	5.84	50.69	5.64	49.06	5.64	−4.63**	1.63
Age at Year 2 follow-up	57.64	5.30	63.21	5.86	62.23	6.15	−5.10**	0.98
Age at Year 3 follow-up	69.24	5.01	74.33	5.98	73.46	6.10	−4.69**	0.87
Months between intake and Year 1	14.31	2.22	13.25	2.84	14.31	2.44	0.53	−1.06
Months between intake and Year 2	27.05	1.91	25.36	1.82	26.85	3.11	0.97	−1.49
Months between intake and Year 3	37.90	2.98	37.13	2.36	38.46	2.30	0.15	−1.33

* $p < 0.05$.
** $p < 0.01$.

school, or in the settings of local non-profit entities (Regional Centers) that contracted with the state to manage services to persons with developmental disabilities. As reported in Howard et al. (2005), Year 1 testing occurred an average of 14.3 months after intake. Thereafter, parents of all children were contacted annually to determine if they were interested in having their children participate in follow-up assessments. Table 1 shows the mean ages of the groups at each assessment and the intervals between assessments. On average, Year 2 testing occurred 23–34 months after intake (M = 27.0 months), and Year 3 assessments occurred 31–43 months after intake (M = 37.9 months).

The examiners selected standardized tests of cognitive skills, language skills, and adaptive behavior that were suited to each child's age and level of functioning. Howard et al. (2005) described the instruments used at intake and at Year 1. After Year 1, adaptive behavior was assessed using the Vineland Adaptive Behavior Scales (VABS). Nonverbal IQ was assessed using the Merrill-Palmer Scales of Development (although the Leiter International Performance Scale was used for one child in the IBT group in Year 3). Full-scale IQ was typically assessed after Year 1 using the developmentally appropriate Wechsler instrument, either the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III or WPPSI-Revised) or the Wechsler Intelligence Scale for Children (WISC-III or WISC-IV). However, one child in the IBT group was administered the Stanford-Binet Intelligence Scale (the 4th edition in Year 2 and the 5th edition in Year 3), and in Year 3 two children in the IBT group were administered the Differential Ability Scales, one IBT child was administered the Slosson Intelligence Test-Revised, and one IBT child was administered the Woodcock-Johnson Tests of Cognitive Abilities III. Receptive and expressive language skills were assessed using a variety of instruments. The most common was the Reynell Developmental Language Scales. Others included the Receptive One-Word Picture Vocabulary Test, the Expressive One-Word Picture Vocabulary Test, the Peabody Picture Vocabulary Test (3rd edition), the Expressive Vocabulary Test, and the Sequenced Inventory of Communication Development-Revised.

Measures for which developmental equivalents were available were converted to developmental quotients (DQs) for analysis using the formula DQ = 100 × developmental equivalent (months)/chronological age (months). When all children are the same age, there is no statistical difference between analyzing standard scores (SSs), developmental equivalents, and DQs. Unlike the other two measures, however, DQs allow valid comparisons to be made among children who have different chronological ages at the same assessment time, and automatically compensate for different intervals between assessment times (cf. Delmolino, 2006; Lord & Schopler, 1989).

2.3.2. Statistical analyses

As in our original study, statistical analyses focused on comparing the mean scores of children in the IBT group with those of children in the AP and GP groups; comparing the mean scores of children in the AP group with those of children in the GP group was of secondary interest. Accordingly, in this study we used the same multiple regression approach we employed in Howard et al. (2005). One term in the regression equation was a contrast that compared mean scores of the children in the IBT group with mean scores of the children in the AP and GP groups, while a second contrast term (orthogonal to the first) compared the mean scores of children in the AP group with the mean scores of children in the GP group. Both contrasts were tested simultaneously, together with two covariates (chronological age at diagnosis and parents' mean years of education) to control for group differences in the covariates.

Separate multiple regression analyses were performed for each of the four assessment times (intake, Year 1, Year 2, and Year 3). Repeated measures analyses examining all four assessment times at once were precluded because not all children were assessed at every follow-up. Restricting the analyses to children with complete assessment records would have eliminated more than half of the children from some analyses. Trends over the 3-year course of treatment were examined by using paired *t*-tests to compare each child's score at one assessment with his or her score at the following assessment.

For every dependent measure, we also determined whether each child achieved a favorable outcome. This was defined as a DQ or SS within the normal range of functioning (i.e., 85 or higher), or a DQ or SS that was at least 15 points (1 standard deviation) higher at the final assessment (Year 2 or Year 3) than at intake. This definition is logically similar to the reliable change index proposed by Jacobson and Truax (1991) for evaluating the effects of treatments. Chi-square tests were used to determine whether the percentage of children with a favorable outcome differed by treatment group, with a separate analysis conducted for each dependent measure.

3. Results

3.1. Ages and assessment times

The assessment chronology for all three groups is summarized in Table 1. Cells in the first five rows include descriptive statistics on chronological ages at diagnosis and at each assessment time. Data in the bottom three rows describe elapsed time between intake and later assessments. Data in the two rightmost columns represent comparisons of group means; asterisks indicate statistically significant differences. These data indicate that, at diagnosis and every subsequent assessment, the average child in the IBT group was younger than the average child in either comparison group; those differences were statistically significant. There was also a statistically significant difference between the mean ages of the AP and GP children at diagnosis, but not at any of the later assessments.

3.2. Analyses of standard scores and developmental quotients

Table 2 presents descriptive statistics and analyses of assessments of cognitive and adaptive skills for each group. Adaptive behavior scores (communication, social, and self-help skills) are expressed as developmental quotients (DQs), while cognitive skills scores and the composite adaptive behavior measure are expressed as standard scores (SSs). For each of the five measures, cells in the first four rows under each group's column list descriptive statistics from each assessment time; results of statistical comparisons of group means at each assessment time are shown in the two rightmost columns. All comparisons controlled for the child's age at diagnosis and the parents' years of education. Asterisks indicate statistical significance. As shown in the two rightmost columns, all Year 1 and Year 2 mean SSs and DQs were significantly higher for the IBT group than for the two comparison groups combined. There were no other statistically significant between-group differences; the mean scores for the IBT group and the two comparison groups combined did not differ significantly at intake, and the mean scores of the AP and GP groups did not differ significantly from each other at intake or at any of the other assessment times on any measure.

The cells in the bottom three rows for each of the five measures in Table 2 summarize changes in mean scores between successive assessments (intake to Year 1, Year 1 to Year 2, and Year 2 to Year 3). Asterisks denote statistically significant improvements (positive values) or declines (negative values) from one year to the next. The IBT group had statistically significant improvements on all measures from intake to Year 1. The AP group had a statistically significant improvement on the cognitive skills SS from intake to Year 1, and statistically significant declines in the self-help DQ and adaptive behavior composite SS from Year 1 to Year 2. The GP group had a statistically significant improvement on the social skills DQ from Year 1 to Year 2. No other changes were statistically significant.

The cells in the penultimate column in the bottom three rows for each measure in Table 2 represent comparisons of the mean change scores of the IBT group and the AP and GP groups combined. Asterisks indicate statistically significant differences in change scores between intake and Year 1 on all measures in favor of the IBT group. The cells in the rightmost

Table 2
Cognitive and adaptive skills scores at intake and Years 1–3, changes between successive assessments, and differences between groups.

Measure	Assessment	IBT group			AP group			GP group			IBT mean minus AP/GP mean	AP mean minus GP mean
		n	*M*	SD	*n*	*M*	SD	*n*	*M*	SD		
Cognitive (SS)	Intake	28	60.57	17.48	16	53.69	13.50	15	61.00	15.27	3.35	−7.31
	Year 1	26	89.88	20.87	16	62.13	19.63	16	69.13	15.13	24.26**	−7.00
	Year 2	22	86.59	20.47	13	59.38	13.38	13	66.08	17.35	23.86**	−6.69
	Year 3	21	89.43	23.99	14	64.43	24.84	13	71.77	19.45	21.47	−7.34
	Intake vs Year 1	25	27.44**	14.18	16	8.44*	15.04	15	8.27	17.20	19.09**	0.17
	Year 1 vs Year 2	20	−2.05	6.89	13	1.77	6.69	13	−1.69	6.79	−2.09	3.46
	Year 2 vs Year 3	17	2.94	9.06	11	1.09	8.62	11	6.55	13.45	−0.88	−5.45
Communication (DQ)	Intake	29	49.23	14.77	16	42.95	15.01	15	47.22	14.39	4.22	−4.27
	Year 1	26	76.47	22.48	16	46.52	21.63	16	52.38	22.34	27.02**	−5.86
	Year 2	22	73.46	28.93	13	45.94	20.50	13	51.26	22.56	24.86*	−5.33
	Year 3	19	72.84	29.46	14	49.40	29.36	13	57.68	22.74	19.45	−8.29
	Intake vs Year 1	26	26.23**	21.27	16	3.57	17.96	15	4.82	21.34	22.05**	−1.25
	Year 1 vs Year 2	20	−0.02	21.08	13	2.22	18.46	13	4.33	15.46	−3.29	-2.11
	Year 2 vs Year 3	16	2.57	21.79	11	4.68	14.68	11	5.68	10.67	−2.61	−1.00
Self-help (DQ)	Intake	29	60.65	16.59	16	56.63	15.54	15	60.30	17.13	2.24	−3.67
	Year 1	26	70.19	16.39	16	62.17	16.49	16	56.39	10.29	10.91*	5.78
	Year 2	22	69.35	16.99	13	52.13	16.46	13	60.68	16.39	12.95**	−8.55
	Year 3	19	67.13	15.78	14	56.11	19.25	13	61.92	13.67	8.22	−5.80
	Intake vs Year 1	26	9.86*	21.24	16	5.54	20.52	15	−3.41	17.93	8.65*	8.95
	Year 1 vs Year 2	20	0.68	17.85	13	−8.53*	14.07	13	5.46	15.26	2.22	−13.99*
	Year 2 vs Year 3	16	−2.66	8.33	11	3.31	18.15	11	1.71	17.01	−5.17	1.60
Social (DQ)	Intake	28	54.43	18.00	16	58.05	23.15	15	56.92	15.50	−3.07	1.13
	Year 1	26	69.47	19.05	16	58.44	28.89	16	49.55	20.21	15.47*	8.89
	Year 2	22	80.66	26.60	13	48.68	25.52	13	61.16	30.09	25.75**	−12.48
	Year 3	19	79.33	27.48	14	52.60	25.56	13	67.85	22.40	19.39	−15.25
	Intake vs Year 1	25	15.21**	22.99	16	0.39	23.86	15	−7.47	23.09	18.62*	7.86
	Year 1 vs Year 2	20	11.85**	17.14	13	−7.93	17.77	13	15.68*	20.45	7.98	−23.61**
	Year 2 vs Year 3[a]	16	3.37	14.81	11	1.57	17.41	11	7.27	19.86	−1.05	−5.70
Composite (SS)	Intake	26	72.00	7.73	16	69.81	10.48	13	69.69	8.54	2.24	0.12
	Year 1	26	81.15	10.95	16	69.25	12.91	16	67.63	10.45	12.72**	1.63
	Year 2	22	79.14	12.06	13	61.08	10.14	13	69.77	12.74	13.71**	−8.69
	Year 3	20	76.00	15.94	14	58.07	15.80	13	65.85	12.23	14.19	−7.77
	Intake vs Year 1	24	8.88**	10.94	16	−0.56	12.04	13	−1.31	11.24	9.77*	0.75
	Year 1 vs Year 2	20	−0.60	10.75	13	−6.54*	8.41	13	4.54	8.29	0.40	−11.08**
	Year 2 vs Year 3	16	−4.31	8.25	11	−4.09	11.23	11	−3.45	7.41	−0.54	−0.64

[a] Age at diagnosis is a significant covariate ($p < 0.05$).
* $p < 0.05$.
** $p < 0.01$.

column in the bottom three rows for each measure represent comparisons of the mean change scores between the AP and GP groups. Statistically significant differences were seen in the changes in self-help DQ, social DQ, and adaptive behavior composite SS from Year 1 to Year 2; in each case, the mean score declined for the AP group but increased for the GP group.

Table 3, which is formatted similarly to Table 2, presents nonverbal IQ, language, and motor skills DQs by group. Those skills were not assessed consistently after Year 1; some children were reassessed only after 2 years of treatment, while others were reassessed only after 3 years of treatment. Accordingly, Table 3 shows statistics for only three points in time: at intake; one year after intake; and either three years after intake or – if no assessment was made at Year 3 – two years after intake. For the IBT group, mean DQs on all measures except motor skills (which were in the normal range at intake) increased significantly from intake to Year 1. Expressive language skills also improved significantly from Year 1 to Year 2/3, while the motor skills mean DQ showed a decline over that same interval. For the AP group, the only statistically significant change was a decline in the mean motor skills DQ from Year 1 to Year 2/3. Over that same interval the GP group's mean expressive language DQ increased significantly, but no other statistically significant changes were found for that group. The penultimate right-hand column of Table 3 again shows several statistically significant differences in change scores favoring the IBT group over the two combined eclectic groups. As indicated in the final column, the only statistically significant difference between the AP and GP groups was on the motor skills DQ assessment at Year 2/3, with children in the AP group scoring lower, on average, than children in the GP group.

The group mean data described in Table 2 are represented graphically in Figs. 2–6, which also show individual scores and the numbers of children in each group whose scores were in the normal range (85 and above) on each measure at each assessment. Data from assessments conducted at intake and Year 1 mirror those reported in our previous paper (Howard et al., 2005): Mean scores on all measures for all three groups were comparable at intake, and the IBT group had significantly higher mean scores one year later than did the AP and GP groups. As discussed above, change scores indicated that, on average, children in the IBT group improved more than children in the other two groups after one year of treatment.

Table 3
Nonverbal IQ, language, and motor skills scores at intake, Year 1, and Year 2 or 3, changes between assessments, and differences between groups.

Measure	Assessment	IBT treatment group			AP treatment group			GP treatment group			IBT mean minus AP/GP mean	AP mean minus GP mean
		n	*M*	SD	*n*	*M*	SD	*n*	*M*	SD		
Non-verbal (DQ)	Intake	20	80.44	12.06	16	67.00	17.13	11	76.65	13.37	9.51	−9.65
	Year 1[b]	24	101.04	18.27	16	73.60	24.79	15	81.08	18.72	23.82**	−7.47
	Year 2/3	24	98.05	17.92	15	69.33	22.08	14	82.20	21.74	22.50**	−12.87
	Intake vs Year 1	20	20.31**	14.97	16	6.61	18.56	11	2.42	13.42	15.41**	4.19
	Year 1 vs Year 2/3	21	−2.20	8.65	15	−2.10	13.63	13	2.78	11.49	−2.36	−4.87
Receptive (DQ)	Intake[b]	29	48.79	20.87	16	45.44	15.23	15	47.29	13.59	2.45	−1.85
	Year 1	26	71.23	21.97	15	51.39	22.44	14	51.95	19.46	19.57*	−0.55
	Year 2/3	25	74.46	25.08	15	49.53	24.61	13	60.31	18.75	19.93*	−10.78
	Intake vs Year 1	26	22.53**	18.31	15	5.27	13.12	13	2.77	11.96	18.42**	2.50
	Year 1 vs Year 2/3	24	2.18	12.39	14	−0.27	20.81	12	4.39	10.49	0.31	−4.66
Expressive (DQ)	Intake	29	49.73	16.34	16	43.90	5.80	15	50.20	12.16	2.78	−6.30
	Year 1	26	69.24	23.20	15	47.31	24.42	14	48.08	14.35	21.56*	−0.77
	Year 2/3	26	83.25	29.88	15	47.98	27.25	13	62.07	23.83	28.73*	−14.10
	Intake vs Year 1	26	20.46**	22.36	15	3.42	22.68	13	−2.84	12.29	19.95*	6.26
	Year 1 vs Year 2/3	24	10.40*	17.11	14	1.51	16.47	12	12.34*	17.75	3.90	−10.82
Motor (DQ)	Intake	28	94.65	17.50	16	89.55	13.09	14	86.96	13.34	6.31	2.59
	Year 1	26	97.30	14.74	16	85.08	12.24	16	85.62	13.62	11.95*	−0.54
	Year 2/3	25	90.17	12.64	12	74.00	13.24	14	86.31	15.83	9.54	−12.30*
	Intake vs Year 1	25	0.63	18.23	16	−4.46	12.82	14	0.83	18.18	2.63	−5.29
	Year 1 vs Year 2/3	23	−8.44*	18.05	12	−9.82*	13.54	14	1.97	20.57	−4.97	−11.78

[b] Mean parental years of education is a significant covariate ($p < 0.05$).
* $p < 0.05$.
** $p < 0.01$.

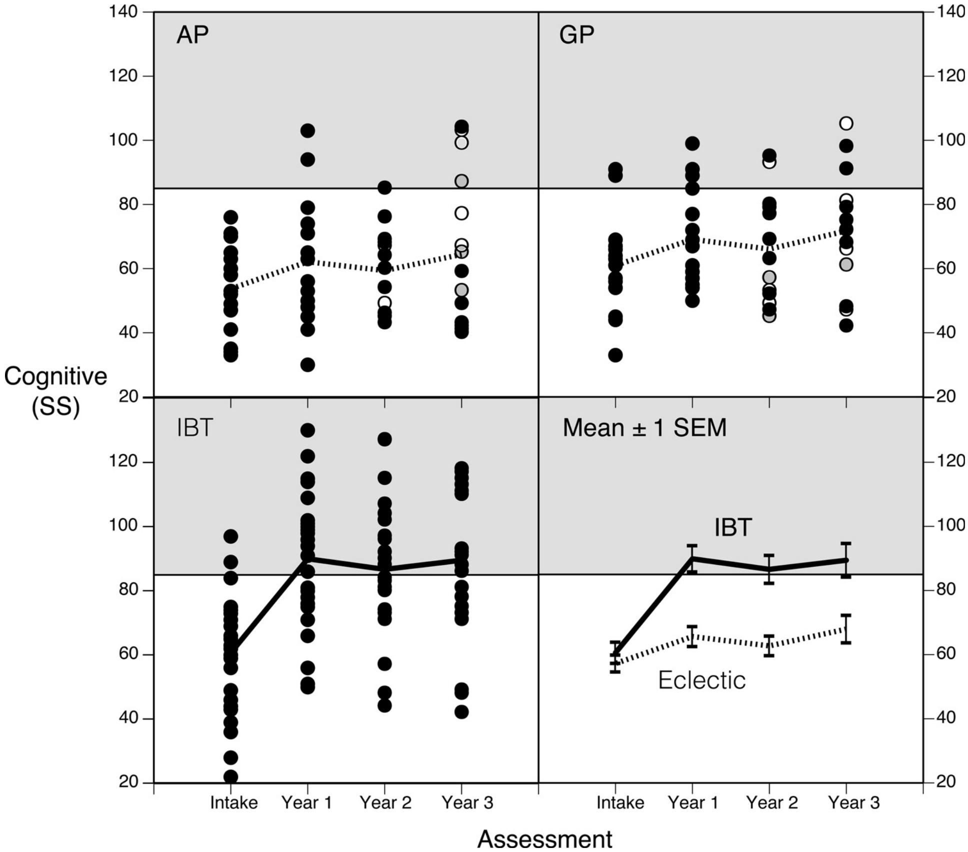

Fig. 2. Cognitive SSs at intake and 1–3 years later. Each dot represents the score for an individual child at that assessment time. Black dots indicate children who received their original treatment at the time of testing; white dots indicate children in the AP group who received GP treatment in the year preceding assessment, or children in the GP group who received AP treatment prior to assessment. Gray dots indicate children whose treatment prior to assessment was not recorded. Scores in the gray region of each panel are in the normal range (85 or higher). The lines in each panel connect the group mean scores at each assessment. The vertical bars in the lower right panel extend ±1 standard error around each group mean.

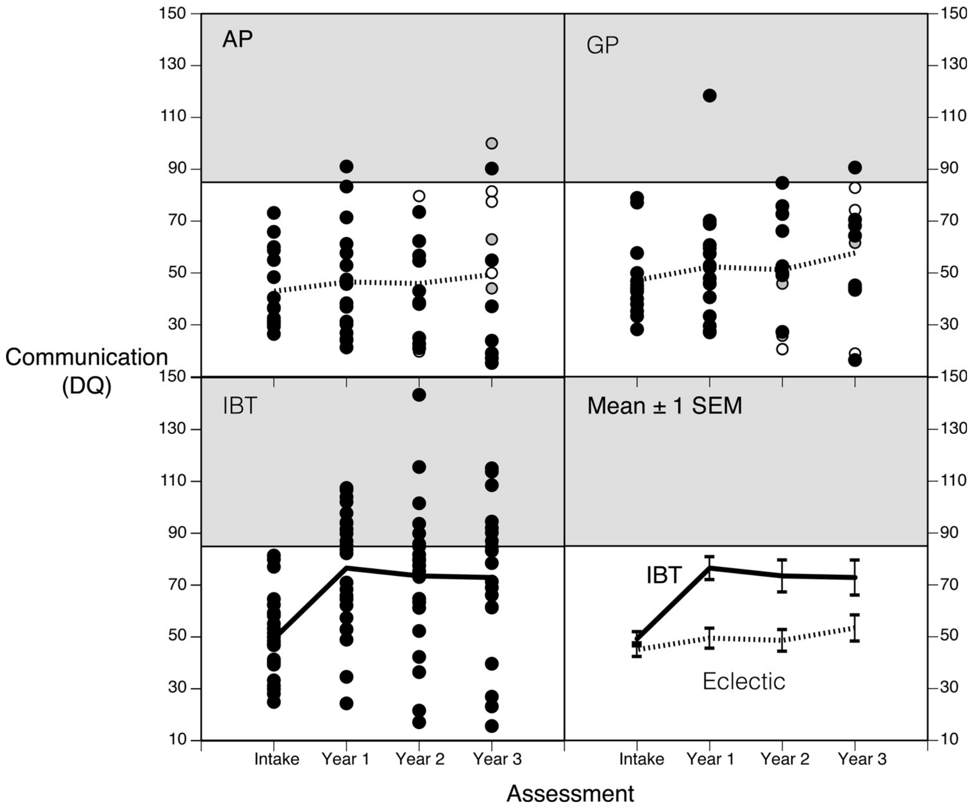

Fig. 3. Communication DQs at intake and 1–3 years later. See Fig. 2 caption for details.

After the first year of treatment, the sharply accelerated trajectory for the IBT group relative to the two other groups did not continue for any measure except the social skills DQ, which increased again from Year 1 to Year 2 before leveling off (Fig. 5). The mean cognitive skills SS for the IBT group remained stable from Year 1 to Year 3 (Fig. 2), while the mean communication skills DQ, self-help skills DQ, and adaptive skills composite SS declined slightly (Figs. 3, 4 and 6, respectively). The mean scores for the GP and AP groups either increased slightly or declined from Year 1 to Year 3 on all measures except social skills DQs, which increased for the GP group (Fig. 5).

In general, the gaps that emerged between the means of the IBT group and the other two groups after one year of treatment remained fairly constant or expanded in favor of the IBT group in Years 2 and 3 (see the lower right-hand panels of Figs. 2–6). Although the mean scores for the children in the IBT group were higher than those of the children in the eclectic treatment group three years after intake, those differences were not statistically significant (see Tables 2 and 3). With one possible exception, that was not because children in the IBT group regressed or because those in the AP and GP groups improved substantially; rather, it was because some children lacked 3-year followup assessments, reducing the Year 3 sample sizes and precluding the detection of statistically significant differences among the group mean scores. The exception was the mean motor skills DQ for the IBT group, which declined slightly from intake to Year 2/3 but remained in the normal range. The AP group's mean motor skills DQs also declined over the course of treatment; that decline was statistically significant and resulted in a Year 3 mean that was below normal (see Table 3).

Given the large improvements in the IBT group after one year of treatment, it may seem surprising that continued treatment did not produce further large gains on most measures; rather, most mean scores remained stable or declined slightly in Years 2 and 3. That finding should be interpreted with caution, however, and in relation to the results for the other two groups. For example, the mean cognitive skills SS for the IBT group was in the normal range after one year of treatment, so further large increases were unlikely. The mean adaptive skills composite SS for the IBT group fell slightly over the course of treatment, but the means for the two other groups fell even more. One plausible explanation for the apparent declines in the mean VABS composite scores is that the programming for these young children emphasized skills other than those assessed by the VABS.

3.3. Analyses of outcomes by type of treatment

Additional analyses were conducted to ascertain the proportions of children in each group who achieved clinically important outcomes by the end of treatment, and the likelihood that each type of treatment would produce such outcomes.

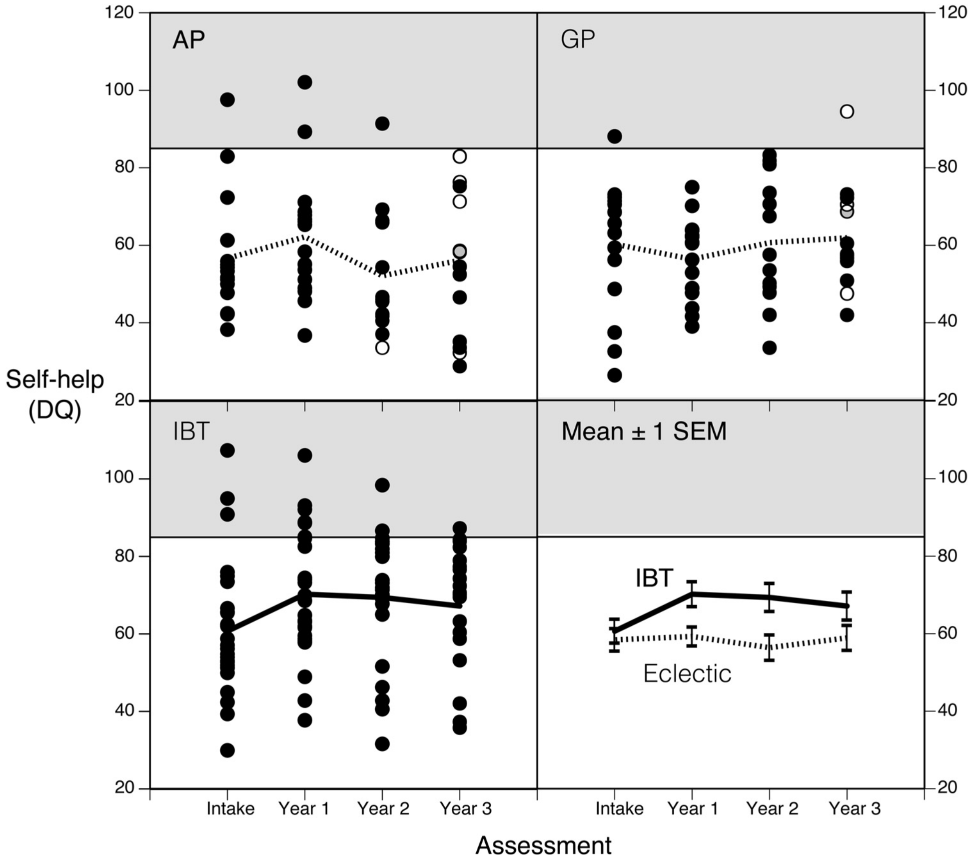

Fig. 4. Self-help DQs at intake and 1–3 years later. See Fig. 2 caption for details.

Table 4 shows the percentage of children in each group who had final (Year 2 or 3) scores in the normal range (i.e., ≥85; third column), final scores that were at least one standard deviation (≥15 points) higher than their intake scores (fifth column), and either of those favorable outcomes (penultimate column). Columns immediately to the right of each of those show odds ratios and probability ratios. To illustrate the odds ratio statistic, consider the data in the fourth column for the cognitive SS. For the IBT group, 61% had a final score ≥85 on that measure; the odds of achieving that favorable outcome were 0.607/(1 − 0.607) = 1.545. For the children in the AP and GP groups combined, 25% had final scores ≥85; the odds of this outcome were 0.250/(1 − 0.250) = 0.333. The ratio of those two odds is 1.545/0.333 = 4.64. This odds ratio of 4.64 is greater than the "neutral" value of 1, indicating that a favorable outcome on the cognitive SS was attained more often by children in the IBT group than by children in the two other groups combined. A likelihood ratio test, which is similar to a chi-square test, confirmed this difference as statistically significant. An odds ratio of 4.64, however, does not signify that children in the IBT group were 4.64 times more likely to have a favorable outcome than children in the AP and GP groups. Such an estimate is better provided by the probability ratio, which is shown in parentheses below each odds ratio in Table 4. The probability ratio for the cognitive SS example is 0.607 (the probability of a final score ≥85 for children in the IBT group) divided by 0.250 (the probability of a final score ≥85 for children in the AP and GP groups combined) = 2.43, indicating that children in the IBT group were 2.43 times more likely to achieve final cognitive SSs in the normal range than were children in the other two groups combined. Probability ratios are more readily interpreted than odds ratios, but statistical tests for group differences utilize odds ratios.

Table 4 shows that the overwhelming majority of the odds ratios and probability ratios favored IBT, indicating that clinically important outcomes as defined here were far more likely to be attained by children who received IBT than by children who received either of the other two treatments. The only exception was that final motor DQ scores were unlikely to be at least one standard deviation above the intake scores. As noted previously, that was likely due to a ceiling effect, in that the mean motor DQ for the IBT group was in the normal range at intake and stayed there over the course of treatment. Double asterisks in Table 4 show that the advantage for IBT children was more likely to be statistically significant when a favorable outcome was defined as a final score ≥85 than when it was defined as an increase of at least 15 points over intake.

Statistically significant differences between the AP and GP groups emerged only for an increase of 15 points or more over intake for social, motor, and adaptive skills composite scores. For those three measures, the odds of a favorable outcome were higher for the GP group than for the AP group. For the cognitive, receptive, and self-help measures, children in the AP group

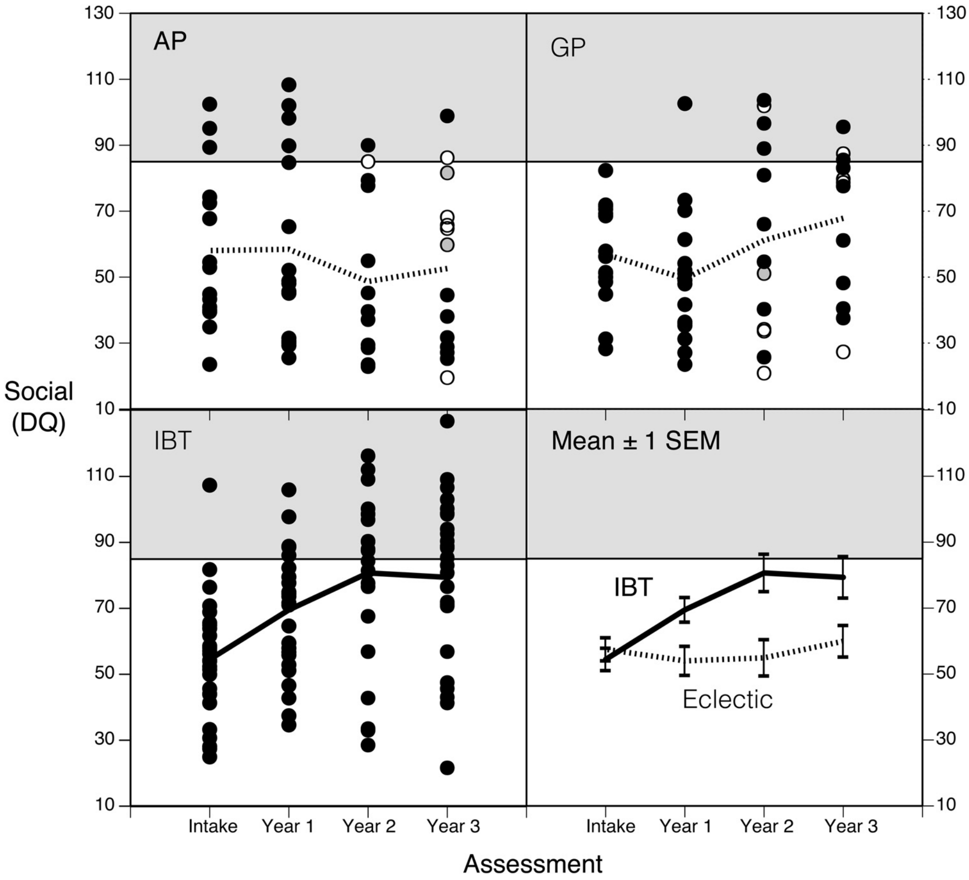

Fig. 5. Social DQs at intake and 1–3 years later. See Fig. 2 caption for details.

were more likely to have favorable outcomes than children in the GP group, though none of those differences was statistically significant. Collectively, these analyses suggest that neither of the comparison treatments was likely to result in favorable outcomes, and that combining the AP and GP groups did not mask any important group differences in outcomes.

Fig. 7 is a graphic representation of the percentages of children in the IBT group and the combined AP and GP groups who had scores in the normal range at each assessment. At intake, those percentages were comparably small for both groups on all measures except the motor skills DQ, on which fairly large proportions of both groups (57% IBT, 47% AP/GP combined) had scores in the normal range. By the end of treatment, a larger percentage of children in the IBT group than in the AP/GP group had scores in the normal range on all measures except the self-help DQ.

Individuals with final scores that were in the normal range ($\geq$85) or at least one standard deviation above intake scores can be readily identified in Fig. 8. In this figure, each child's score on each measure is plotted as a function of his or her score at intake (on the *x*-axis) and the change from intake to the final assessment (on the *y*-axis; the final assessment was made at Year 2 if the child was not assessed at Year 3). Final scores in the normal range appear in the dark gray region of each panel, and scores representing increases of at least one standard deviation over intake are in the light gray regions. Both regions are populated by more children in the IBT group (closed circles) than by children in the other two groups (open symbols). That is, more of the children who received IBT had final outcomes that constituted clinically important changes over baseline than did children who received either of the other two treatments.

An important question is whether children in this study who attained normal levels of functioning at any point maintained those levels over the course of treatment. That question is difficult to answer, because only a portion of the children in each group had scores in the normal range at any assessment time, and not all children were assessed at both Year 2 and Year 3. Nevertheless, the question is sufficiently important to merit an attempt to answer it. For this analysis, children were classified into four categories of outcomes: (a) scored <85 one year and remained <85 the next year; (b) scored <85 one year but scored $\geq$85 the next year (i.e., transitioned to a normal range of functioning); (c) scored $\geq$85 one year but scored <85 the next year (i.e., regressed); and (d) scored $\geq$85 one year and remained $\geq$85 the following year. Those categories were then combined across measures to calculate the probability of each of the four outcomes for each year-to-year assessment transition. Separate calculations were made for children in the IBT group and for children in the combined AP and GP groups.

Results of these analyses are illustrated by the Sankey diagram shown in Fig. 9. In this figure, arrows are not just proportional in width to the quantities the represent; they are also horizontal if they represent children who maintained

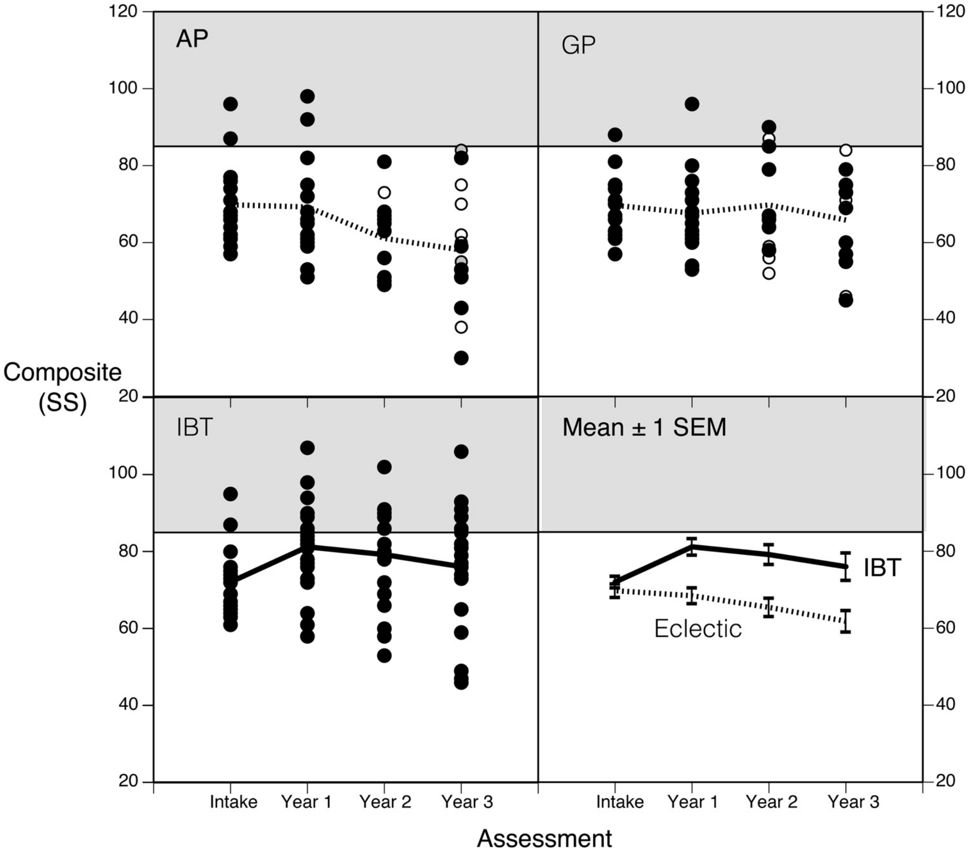

Fig. 6. Composite adaptive skills SSs at intake and 1–3 years later. See Fig. 2 caption for details.

assessed levels of functioning, they slant upward for children who improved, and they slant downward for children who regressed from one year to the next. The figure should be interpreted cautiously, because it represents data that were collapsed across measures and is based upon other suboptimal manipulations. Nevertheless, several intriguing trends are suggested. One is that most children who moved from below-normal to normal-range functioning did so after one year of treatment. For both groups, the probability of moving into the normal range was higher from intake to Year 1 than from Year 1 to Year 2, or from Year 2 to Year 3 (indicated by the upward-slanting arrows in Fig. 9). Stated differently, the prospect of achieving scores in the normal range diminished with each additional year of treatment, but the likelihood of scoring in the normal range was substantially and consistently higher for children in the IBT group than for children in the AP/GP groups combined at all three years post-intake (as shown by the percentages in the upward-slanting arrows). For children in the AP/GP group, if a score $\geq$85 was not attained after one year of treatment, the prospects for attaining a normal score were extremely dim.

A second general trend, confirming analyses presented in preceding tables and figures, is that children in the IBT group were far more likely to score in the normal range at all three post-intake assessments than were children in the two comparison groups. Further, percentages shown in the upward slanting arrows indicate that children in the IBT group were more than three times as likely as children in the AP and GP groups to have scores that moved them from the below-normal to the normal range at Years 1–3. That advantage was not limited to Year 1 scores; it remained relatively consistent throughout all three years of the study.

A final trend, illustrated by the downward slanting arrows in Fig. 9, is that regressions from normal to below-normal range scores were much more common for children in the AP/GP group than for children in the IBT group. In fact, children in the AP/GP group were 3.45 times as likely to regress as to advance during the first year of treatment, 4.45 times more likely to regress than advance during the second year of treatment, and 4.91 times more likely to regress than advance in the third year of treatment. The opposite pattern was seen for children in the IBT group, where advancements were 2.48 times as likely as regressions during the first year of treatment. Advancements and regressions occurred about equally often between Year 1 and Year 2 for the IBT group (the ratio was 1.08 in favor of advancements), but in the third year of treatment advancements were 1.75 times as frequent as regressions. Collectively, these findings suggest that children who received IBT were much more likely to attain and maintain normal levels of functioning than were children who received either of the other treatments.

Table 4
Percent of children with favorable outcomes, and odds ratios and probability ratios for each measure.

Measure	Group	Final score ≥ 85	Odds ratio (probability ratio)	Final score ≥ 15 points above intake	Odds ratio (probability ratio)	Either desirable outcome	Odds ratio (probability ratio)
Cognitive (SS)	IBT	61% ($n=28$)	4.64** (2.43)	81% ($n=27$)	8.00** (2.30)	82% ($n=28$)	8.78** (2.39)
	AP/GP combined	25% ($n=32$)		35% ($n=31$)		34% ($n=32$)	
	AP	25% ($n=16$)	1.00 (1.00)	38% ($n=16$)	1.20 (1.13)	38% ($n=16$)	1.32 (1.20)
	GP	25% ($n=16$)		33% ($n=15$)		31% ($n=16$)	
Non-verbal (DQ)	IBT	85% ($n=27$)	8.40** (2.10)	60% ($n=20$)	3.00 (1.80)	85% ($n=27$)	6.52** (1.82)
	AP/GP combined	41% ($n=32$)		33% ($n=27$)		47% ($n=32$)	
	AP	31% ($n=16$)	0.45 (0.63)	31% ($n=16$)	0.80 (0.86)	44% ($n=16$)	0.78 (0.88)
	GP	50% ($n=16$)		36% ($n=11$)		50% ($n=16$)	
Receptive (DQ)	IBT	26% ($n=27$)	5.08* (4.02)	78% ($n=27$)	8.17** (2.59)	85% ($n=27$)	10.45** (2.40)
	AP/GP combined	6% ($n=31$)		30% ($n=30$)		35% ($n=31$)	
	AP	6% ($n=16$)	0.93 (0.94)	31% ($n=16$)	1.14 (1.09)	38% ($n=16$)	1.20 (1.13)
	GP	7% ($n=15$)		29% ($n=14$)		33% ($n=15$)	
Expressive (DQ)	IBT	46% ($n=28$)	5.85** (3.60)	82% ($n=28$)	9.20** (2.46)	82% ($n=28$)	9.66** (2.55)
	AP/GP combined	13% ($n=31$)		33% ($n=30$)		32% ($n=31$)	
	AP	13% ($n=16$)	0.93 (0.94)	31% ($n=16$)	0.82 (0.88)	31% ($n=16$)	0.91 (0.94)
	GP	13% ($n=15$)		36% ($n=14$)		33% ($n=15$)	
Commun-ication (DQ)	IBT	36% ($n=28$)	3.89* (2.86)	68% ($n=28$)	2.25 (1.40)	75% ($n=28$)	3.00* (1.50)
	AP/GP combined	13% ($n=32$)		48% ($n=31$)		50% ($n=32$)	
	AP	13% ($n=16$)	1.00 (1.00)	38% ($n=16$)	0.40 (0.63)	38% ($n=16$)	0.36 (0.60)
	GP	13% ($n=16$)		60% ($n=15$)		63% ($n=16$)	
Self-help (DQ)	IBT	11% ($n=28$)	3.72 (3.43)	39% ($n=28$)	1.36 (1.22)	43% ($n=28$)	1.65 (1.37)
	AP/GP combined	3% ($n=32$)		32% ($n=31$)		31% ($n=32$)	
	AP	0% ($n=16$)	0.00 (0.00)	38% ($n=16$)	1.65 (1.41)	38% ($n=16$)	1.80 (1.50)
	GP	6% ($n=16$)		27% ($n=15$)		25% ($n=16$)	
Social (DQ)	IBT	54% ($n=28$)	4.12* (2.45)	67% ($n=27$)	2.77 (1.59)	71% ($n=28$)	3.65* (1.76)
	AP/GP combined	22% ($n=32$)		42% ($n=31$)		41% ($n=32$)	
	AP	13% ($n=16$)	0.31 (0.40)	25% ($n=16$)	0.22* (0.42)	25% ($n=16$)	0.26 (0.44)
	GP	31% ($n=16$)		60% ($n=15$)		56% ($n=16$)	
Motor (DQ)	IBT	57% ($n=28$)	1.51 (1.22)	19% ($n=27$)	0.91 (0.93)	57% ($n=28$)	1.51 (1.22)
	AP/GP combined	47% ($n=32$)		20% ($n=30$)		47% ($n=32$)	
	AP	31% ($n=16$)	0.27 (0.50)	0% ($n=16$)	0.00** (0.00)	31% ($n=16$)	0.27 (0.50)
	GP	63% ($n=16$)		43% ($n=14$)		63% ($n=16$)	
Composite (SS)	IBT	36% ($n=28$)	8.33** (5.71)	16% ($n=25$)	1.65 (1.55)	36% ($n=28$)	5.37* (3.81)
	AP/GP combined	6% ($n=32$)		10% ($n=29$)		9% ($n=32$)	
	AP	0% ($n=16$)	0.00 (0.00)	0% ($n=16$)	0.00* (0.00)	0% ($n=16$)	0.00* (0.00)
	GP	13% ($n=16$)		23% ($n=13$)		19% ($n=16$)	

* Odds ratio differs significantly from 1 ($p < 0.05$).
** Odds ratio differs significantly from 1 ($p < 0.01$).

4. Discussion

4.1. Differential treatment outcomes

Our 2005 study evaluated outcomes for 61 children with autism who received just over one year of either IBT or one of two eclectic interventions. Although the three groups were similar at intake, children who received IBT had significantly higher mean scores after one year of treatment than those who received eclectic interventions. The present study extended

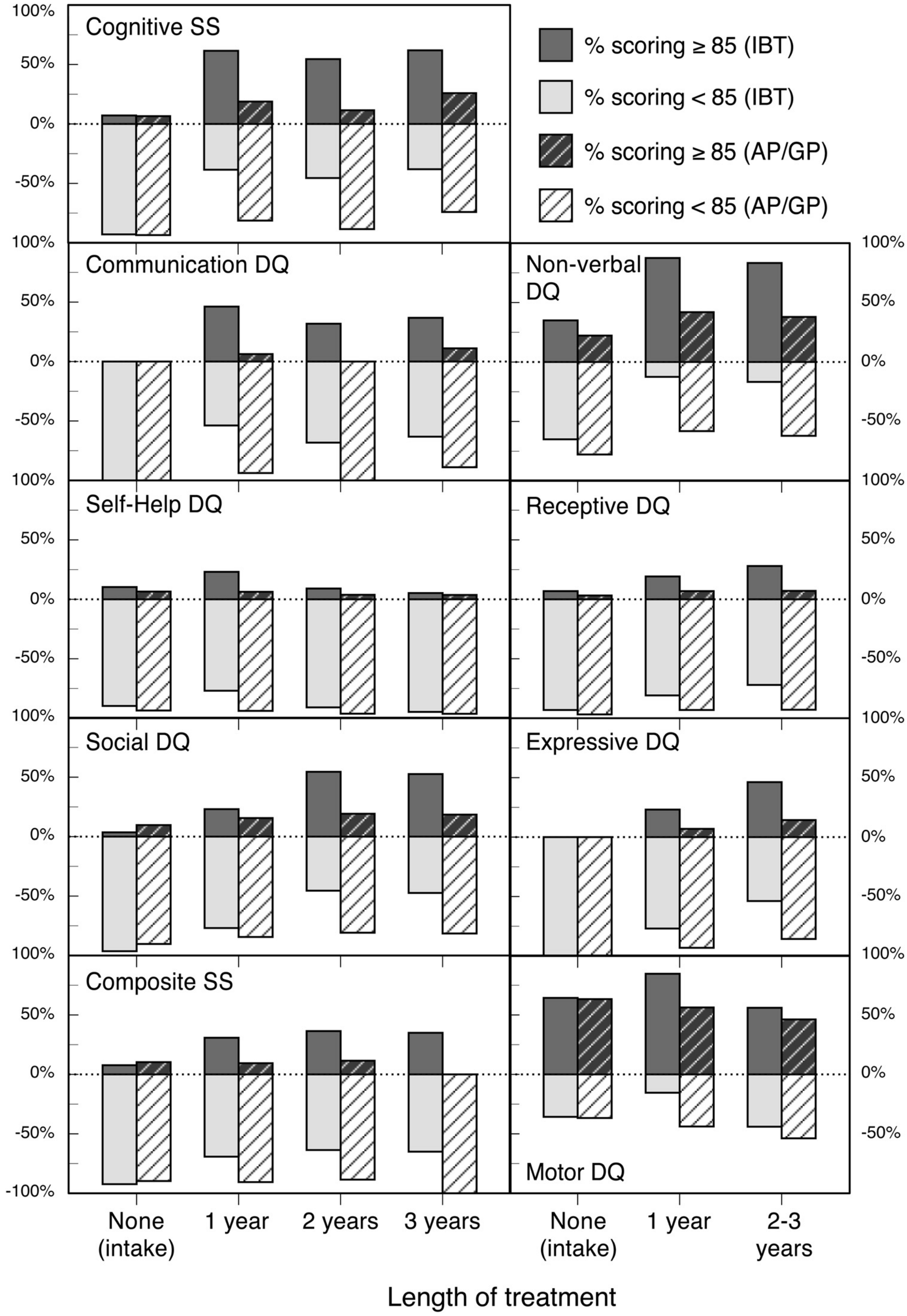

Fig. 7. Percent of children in each treatment group with a score in the normal range (SS or DQ ≥85) at intake and 1–3 years after intake.

those findings by showing that the largest gains generally occurred in the first year of treatment and in IBT children only, and that the advantage experienced by IBT children after one year of treatment was maintained throughout the second and third years of treatment. Indeed, three years after treatment began, mean scores on standardized assessments of cognitive, language, adaptive, and motor skills were higher for children in the IBT group than they were for children in the eclectic intervention groups.

At their final assessment, 61% of the children who received IBT tested within the average range of cognitive functioning, compared with only 25% of the children who received eclectic treatment. That is, children in the IBT group were more than twice as likely to attain a cognitive skills score in the normal range as children in the two eclectic intervention groups. Final

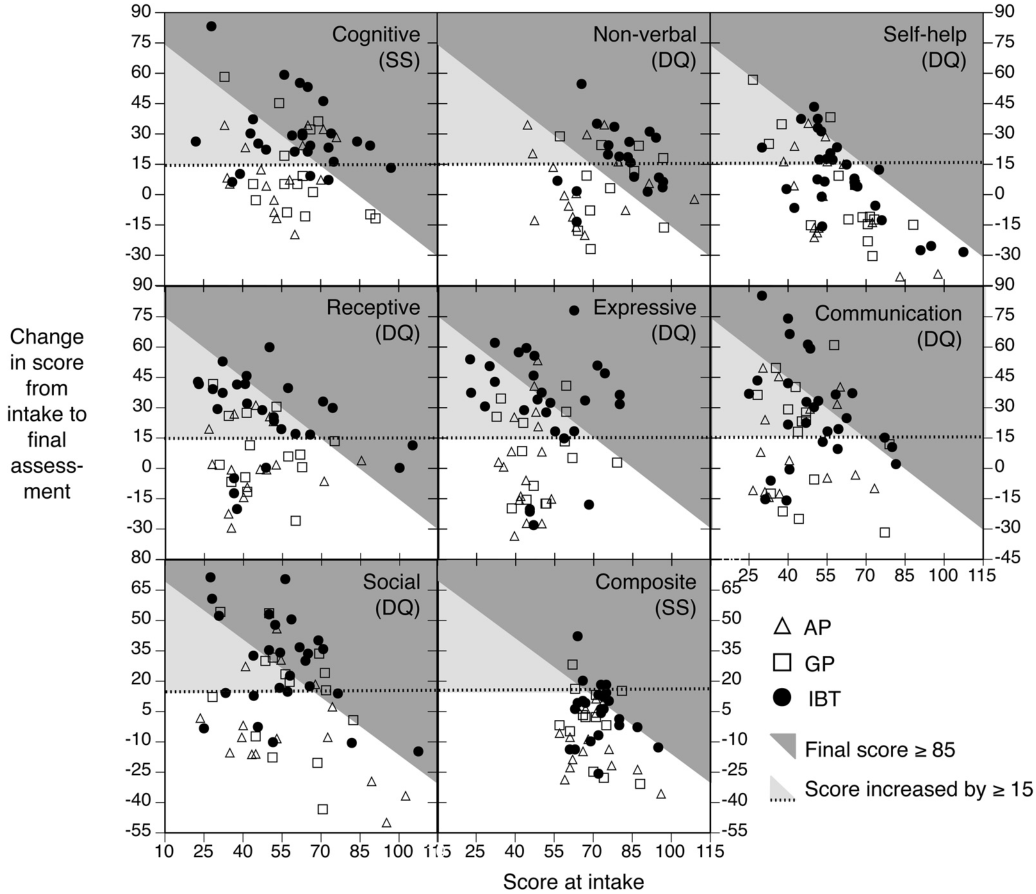

Fig. 8. Scores for individual children on each measure, plotted as a function of the value at intake along the *x*-axis, and the change from intake to Year 3 (or Year 2 if the child was not assessed at Year 3) along the *y*-axis. Scores of children in the IBT group are shown as solid circles, scores of children in the AP group are represented by open triangles, and scores of children in the GP group are shown as open squares. Final scores in the normal range (≥85) appear in the dark gray region of each panel, and final scores <85 but at least 15 points higher than at intake (i.e., above the dotted line in each panel) appear in the light gray region of each panel.

assessment scores on other measures showed similar patterns. Compared to children who received eclectic interventions, children who received IBT were twice as likely to score in the normal range on the final assessment of nonverbal skills, approximately three times as likely to score in the normal range on the final assessments of communication and adaptive skills, approximately four times as likely to score within the normal range on the final assessments of receptive and expressive communication skills, and almost six times more likely to have a final adaptive behavior skills composite score within the normal range.

As they were at Year 1, average outcomes at Years 2 and 3 were worse for children in the AP and GP groups than for children in the IBT group, while average outcomes for the two eclectic intervention groups did not differ significantly from each other. The mean score for the GP group was higher than the mean score for the AP group on some measures in some years, but there were no statistically reliable differences between outcomes produced by the two eclectic treatments. Additionally, both eclectic treatments performed substantially worse than IBT in producing standardized test scores in the normal range of functioning, and neither eclectic treatment was more likely than the other to produce a favorable outcome. The results for the AP intervention might be surprising to some readers because that intervention was intensive and designed specifically for children with autism. Despite these features, no child from the AP group scored in the normal range on the final assessment of adaptive functioning. In contrast, more than one-third of the children in the IBT group achieved a normal-range score on the final assessment of adaptive skills. These findings are especially important given the critical contribution of adaptive skills to independent functioning throughout the lifespan.

Although scores in the normal range are certainly desirable outcomes, so are other clinically significant improvements. Changes in test scores that do not reach the normal range may nonetheless reflect the acquisition of many skills that enhance independent functioning, which in turn produces economic savings due to reduced need for specialized services (Jacobson, Mulick, & Green, 1998; Motiwala, Gupta, Lilly, Ungar, & Coyte, 2006). About one-third of the children in this study who received AP or GP interventions had final scores on tests of cognitive or adaptive skills that were at least 15 points higher than

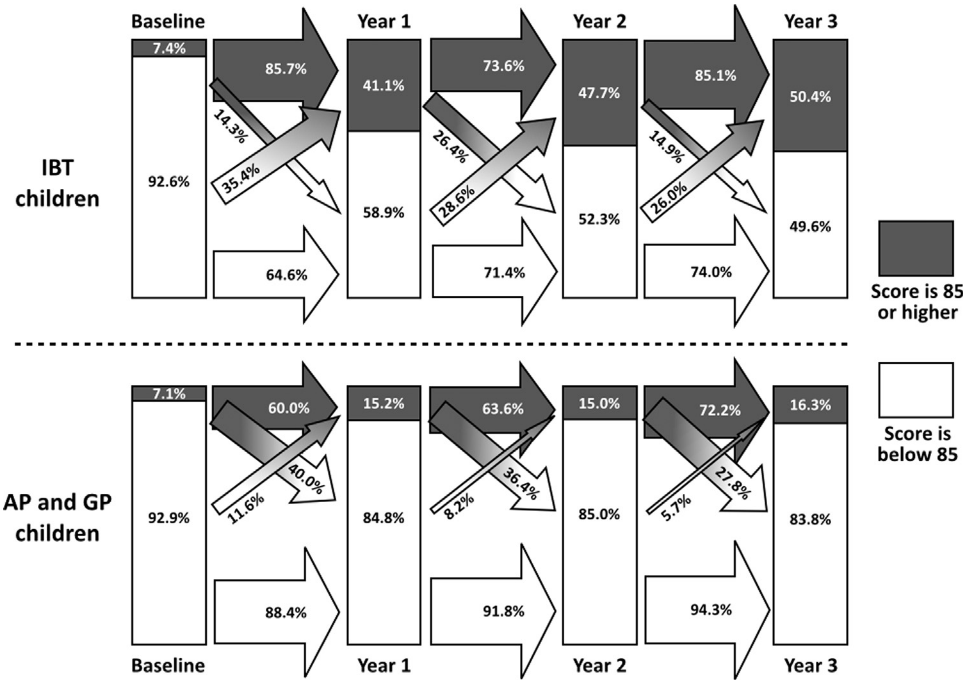

Fig. 9. Percentages of children who scored in the normal range (gray region) or below 85 (white region) at each assessment time, and who transitioned from one of those ranges to another on successive assessments. Horizontal arrows indicate maintenance of scores in the normal range (gray arrows) or in the below-normal range (white arrows). Upward-slanting arrows indicate changes from below-normal to normal-range scores, and downward-slanting arrows represent changes from normal to below-normal range scores.

their intake scores, suggesting that those interventions may produce some benefit for some children with autism. Children in the IBT group, however, were more than twice as likely as children in the other two groups to show changes of that magnitude over the course of treatment. Differences on most other measures were somewhat smaller but equally clear and in the same direction. Motor skills scores were an exception, as they were somewhat more likely to increase by at least 15 points among children in the AP and GP groups than among children in the IBT group. However, that difference was not statistically significant.

The multiple regression approach we used for most of our statistical analyses accommodated individual differences (e.g., in parental education and age at diagnosis), but of course those analyses focused on group data. Group comparisons are appropriate for determining which of two or more treatments is generally most effective; however, we urge caution in relying exclusively on group statistics to prognosticate about individual children. It is clear from the individual data presented here (Figs. 2–6) that not all children within each treatment group responded similarly to that treatment. Research correlating child characteristics with differential outcomes might help identify categories of children who are more or less likely to respond well to a given treatment on average, but more precise information about the effects of treatments on individuals with varying characteristics could be gleaned from studies using single-case research designs, perhaps in combination with elements of between-groups designs (Green, 2008; Guyatt et al., 2008; Larson, 1990; Morgan & Morgan, 2001; Powers et al., 2006). Research methods that focus on changes in individual behavior with treatment could also enable analyses of the differential effectiveness of elements of multicomponent treatments like IBT (e.g., Heyvaert, Maes, Van den Noortgate, Kuppens, & Onghena, 2012) as well as treatment targets that function as behavioral cusps to bring the individual's behavior into contact with new contingencies of reinforcement, thereby producing even more widespread behavior change (Rosales-Ruiz & Baer, 1997).

4.2. Changes over the course of treatment

In this study, the changes that occurred during the first year of treatment were generally maintained throughout the second and third years for children in all three groups. Group mean scores in Years 2 and 3 tended to remain within ±5 points of the corresponding group means at the end of Year 1, with the large differences in favor of IBT after one year largely persisting throughout Years 2 and 3. Other studies comparing IBT with eclectic treatment over similar time periods have produced similar findings (Cohen et al., 2006; Eikeseth et al., 2007). One difference is that the IBT advantage was larger after a mean of 31.4 months of treatment than after one year of treatment in the study by Eikeseth et al. (2007). That may be related to the fact that the children studied by Eikeseth et al. were older when they started treatment than the children in our study and the study by Cohen et al.

(2006), but it might also reflect differences in other child characteristics or the treatment packages (e.g., variations in targets, priorities, procedures, etc.).

Measures of some skill domains in our study deviated from the trends just described. For instance, the mean motor and self-help scores for the IBT group were higher than those for either eclectic intervention group at the end of Year 1, but the differences between the final group means on those measures were not statistically significant. That was at least partly due to reduced sample sizes at Year 3. It should also be reiterated that motor skills were not delayed substantially for any of the groups at intake, and motor and self-help skills were not among the highest priority treatment targets for many of the children who received IBT.

The fact that most of the largest improvements in the IBT group occurred after one year of treatment might lead some to conclude that there is little benefit in extending treatment beyond the first year. Such a conclusion might be warranted if there were compelling evidence to support predictions that improvements would persist if treatment were to end after one year. Our study cannot speak to that hypothesis, because none of the children in the IBT group received just one year of treatment. Nor are we aware of other studies that have tested that hypothesis directly. One group of researchers did, however, evaluate the performances of 23 young children with autism two years after they had completed a 2-year course of IBT (Kovshoff, Hastings, & Remington, 2011). They found that a subgroup of 9 children who had statistically significant increases on tests of cognitive and adaptive skills during treatment maintained those gains after two years with no treatment, but the scores of the other 14 children decreased significantly. Analyses showed that the first subgroup had higher baseline scores and received more intensive treatment than did the second subgroup. Although limited, those findings corroborate our clinical observations that terminating IBT prematurely can be detrimental to many children with autism.

Ending IBT after one year might also be justified if it were reasonably certain that extending treatment would be unlikely to produce further clinically significant gains. Again, we have found no compelling evidence to support that prediction. On the contrary, some children in our IBT group made marked improvements in Years 2 and 3 (e.g., see the upward-pointing arrows in Fig. 9). Other researchers have also documented meaningful improvements occurring in the second, third, and fourth year of IBT (e.g., Cohen et al., 2006; Eikeseth et al., 2007; Sallows & Graupner, 2005). We speculate that given the pervasive and substantial skill deficits exhibited by many young children with autism, one and even two years of IBT is not likely to produce gains that will persist over long periods of time without specialized intervention. The first 1–2 years of IBT are typically focused on building many basic, foundational skills. Further intensive treatment seems essential for solidifying those repertoires and for building the more complex social, language, and academic skills required to function successfully in regular school and community settings.

4.3. Limitations

Participants in this study were not randomly assigned to groups; instead, treatment assignments primarily reflected parental preferences and education team decisions. In Howard et al. (2005), however, we demonstrated empirically that the three groups were functionally equivalent at intake. The only statistically significant group differences were in parental education (parents of children in the IBT group averaged one year more of education than parents of children in the AP and GP groups) and age at diagnosis (children in the IBT group were diagnosed an average of 5 months earlier than children in the GP group, who in turn were diagnosed an average of 4 months earlier than children in the AP group). Both variables were controlled for statistically in subsequent data analyses, though control was rarely necessary because individual scores almost never covaried with parental education or age at diagnosis.

Another limitation is that some children switched between the AP and GP treatments during Years 2 and 3. We have no information about the reasons for those shifts, but it would be unusual for an education team to recommend moving a child out of an effective program and for the child's family to approve such a change. Therefore, we speculate that the changes may speak to the lack of efficacy of either eclectic approach. The data showed that neither eclectic treatment reliably produced meaningful benefits, and when children switched from one eclectic treatment to the other, there was rarely any improvement with the new treatment. These findings imply that the two eclectic treatments were essentially indistinguishable in their efficacy, and that our analyses and conclusions were not compromised by the fact that some children switched from one eclectic treatment to the other.

The impact of mortality on our findings should be considered. Virtually all children were assessed in all domains at intake and Year 1, but participation rates were lower in subsequent years. The reduced sample sizes forced us to combine data from Years 2 and 3 to analyze outcomes for the nonverbal intelligence, receptive language, expressive language, and motor skills measures. That precluded mapping developmental trajectories for those domains as precisely as we did for other domains. It is important to note, however, that mortality does not seem to have biased the overall findings. In fact, imputation analyses suggest that the group differences we observed were not artifacts of mortality; if anything, the advantage of IBT over the eclectic treatments would likely have been greater if more comprehensive assessment data were available for Years 2 and 3.

The primary limitation of our study may be that there were no measures of the integrity with which any of the treatments was delivered, as we reported in Howard et al. (2005). Additionally, each treatment comprised a number of components, and it was not feasible to parse out the contributions of individual components to the outcomes. Nonetheless, our findings converge with those of other studies in which IBT and a comparison eclectic treatment program had similar elements, intensity, and duration (e.g., Cohen et al., 2006; Eikeseth et al., 2007). They add to the growing body of evidence that IBT produces significantly larger increases on standardized measures of cognitive and adaptive functioning than other treatments. Although those measures do not capture all repertoires that may be influenced by intervention, they are considered more objective than indices like classroom placement, and correlate positively with other measures of overall and long-term functioning. Thus, there is general consensus among autism researchers that protocols for evaluating treatment effects must include certain standardized instruments (e.g., Eldevik et al., 2009, 2010; Fein et al., 2013; Martin, Bibby, Mudford, & Eikseth, 2003; Mundy, 1993; Wolery & Garfinkle, 2002), Collectively, this study and others that used such protocols clearly indicate that IBT is an effective, evidence-based treatment for young children diagnosed with autism.

Acknowledgements

This study was supported in part by Valley Mountain Regional Center and California State University, Stanislaus. We also wish to express our appreciation to Jessica Bailey for her assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ridd.2014.08.021.

References

Bondy, A. S., & Frost, L. A. (1994). *The picture exchange communication system: Training manual*. Cherry Hill, NJ: Pyramid.
Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Developmental and Behavioral Pediatrics, 27*, S145–S155.
Delmolino, L. M. (2006). Brief report: Use of DQ for estimating cognitive ability in young children with autism. *Journal of Autism and Developmental Disorders, 36*, 959–963.
Eikeseth, S. (2009). Outcome of comprehensive psycho-educational interventions for young children with autism. *Research in Developmental Disabilities, 30*, 158–178.
Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism: A 1-year comparison controlled study. *Behavior Modification, 2002*, 49–68.
Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification, 31*, 264–278.
Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders, 36*, 211–224.
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*, 439–450.
Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2010). Using participant data to extend the evidence base for intensive behavioral intervention for children with autism. *American Journal on Intellectual and Developmental Disabilities, 115*, 381–405.
Eldevik, S., Hastings, R. P., Jahr, E., & Hughes, J. C. (2012). Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *Journal of Autism and Developmental Disorders, 42*, 210–220.
Fein, D., Barton, M., Eigsti, I., Kelley, E., Naigles, L., Schultz, R. T., & Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry, 54*, 195–205.
Fischer, J. L., Howard, J. S., Sparkman, C. R., & Moore, A. G. (2009). Establishing generalized syntactical responding in young children with autism. *Research in Autism Spectrum Disorders, 4*, 76–88.
Green, G. (2008). Single-case research methods for evaluating treatments for ASD. In S. C. Luce, D. S. Mandell, C. Mazefsky, & W. Seibert (Eds.), *Autism in Pennsylvania: A symposium issue of the Speaker's Journal of Pennsylvania Policy* (pp. 119–132). Harrisburg, PA: Legislative Office for Research Liaison, Pennsylvania House of Representatives.
Green, G. (2011). Early intensive behavior analytic intervention for autism spectrum disorders. In E. Mayville & J. Mulick (Eds.), *Behavioral foundations of effective autism treatment* (pp. 183–199). Cornwall-on-Hudson, NY: Sloan Publishing.
Green, G., Brennan, L. C., & Fein, D. (2002). Intensive behavioral treatment for a toddler at high risk for autism. *Behavior Modification, 26*, 69–102.
Guyatt, G., Rennie, D., Meade, M., & Cook, D. J. (2008). *Users' guides to the medical literature: A manual for evidence-based clinical practice* (2nd ed.). New York: McGraw-Hill Professional.
Heyvaert, M., Maes, B., Van den Noortgate, W., Kuppens, S., & Onghena, P. (2012). A multilevel meta-analysis of single-case and small-*n* research on interventions for reducing challenging behavior in persons with intellectual disabilities. *Research in Developmental Disabilities, 33*, 766–780.
Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359–383.
Jacobson, J. W., Mulick, J. A., & Green, G. (1998). Cost-benefit estimates for early intensive behavioral intervention for young children with autism – General model and single state case. *Behavioral Interventions, 13*, 201–226.
Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
Kovshoff, H., Hastings, R., & Remington, B. (2011). Two-year outcomes for children with autism after the cessation of early intensive behavioral intervention. *Behavior Modification, 35*, 427–450.
Larson, E. B. (1990). N-of-1 clinical trials: A technique for improving medical therapeutics. *Western Journal of Medicine, 152*, 52–56.
Lord, C., & Schopler, E. (1989). The role of age at assessment, developmental level, and test in the stability of intelligence scores in young autistic children. *Journal of Autism and Developmental Disorders, 19*, 483–499.
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.
Martin, N., Bibby, P., Mudford, O. C., & Eikseth, S. (2003). Toward the use of a standardized assessment for young children with autism: Current assessment practices in the UK. *Autism, 7*, 321–330.
Morgan, D. L., & Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist, 56*, 119–127.
Motiwala, S. S., Gupta, S., Lilly, M. B., Ungar, W. J., & Coyte, P. C. (2006). The cost-effectiveness of expanding intensive behavioural intervention to all autistic children in Ontario. *Healthcare Policy, 1*(2), 135–151.
Mundy, P. (1993). Normal versus high-functioning status in children with autism. *American Journal on Mental Retardation, 97*, 381–384.
National Autism Center (2009). *National standards report*. Randolph, MA: National Autism Center.

Powers, S. C., Piazza-Waggoner, C., Jones, J. S., Ferguson, K. S., Daines, C., & Acton, J. D. (2006). Examining clinical trial results with single-subject analysis: An example involving behavioral and nutrition treatment for young children with cystic fibrosis. *Journal of Pediatric Psychology, 31*, 574–581.
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.
Remington, B., Hastings, R. P., Kovshoff, H., Espinosa, F., Jahr, E., Brown, T., & Ward, N. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation, 112*, 418–438.
Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology, 37*, 8–38.
Rosales-Ruiz, J., & Baer, D. M. (1997). Behavioral cusps: A developmental and pragmatic construct for behavioral analysis. *Journal of Applied Behavior Analysis, 30*, 533–544.
Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*, 417–438.
Schmidt, M. (2008). The Sankey diagram in energy and material flow management Part I: History. *Journal of Industrial Ecology, 12*, 82–94.
Smith, T., Groen, A. D., & Wynne, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation, 105*, 269–285.
Wolery, M., & Garfinkle, A. N. (2002). Measures in intervention research with young children who have autism. *Journal of Autism and Developmental Disorders, 32*, 463–478.
Zachor, D. A., Ben-Itzchak, E., Rabinovich, A., & Lahat, E. (2007). Change in autism core symptoms with intervention. *Research in Autism Spectrum Disorders, 1*, 304–307.