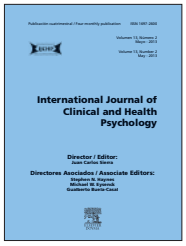
International Journal of Clinical and Health Psychology (2013) **13**, 91–100

ELSEVIER DOYMA

International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

Prediction of treatment outcomes and longitudinal analysis in children with autism undergoing intensive behavioral intervention

Javier Virues-Ortega*[,a], Víctor Rodríguez[b], C.T. Yu[a]

[a]*University of Manitoba and St. Amant Research Centre, Canada*
[b]*Fundación Planeta Imaginario, Spain*

Received September 4, 2012; accepted March 18, 2013

KEYWORDS
Autism;
Predictors;
Applied behavior analysis;
Quasi-experiment (interrupted time-series with one group)

Abstract Outcome prediction is an important component of treatment planning and prognosis. However, reliable predictors of intensive behavioral intervention (IBI) have not been clearly established. IBI is an evidence-based approach to the systematic teaching of academic, social, verbal, and daily living skills to individuals with autism spectrum disorder. Incorporating longitudinal analysis to IBI outcome studies may help to identify outcome predictors of clinical value. Twenty-four children with autism underwent on average two years of IBI and completed language, daily living skills, cognitive, and motor assessments (Early Learning Accomplishment Profile and the Learning Accomplishment Profile-Diagnostic, 3rd edition) every six months. We used multilevel analysis to identify potential longitudinal predictors including gender, age, intervention intensity, intervention duration, total intervention time, and pre-intervention functioning. Results indicated that total intervention time, pre-intervention functioning, and age caused the greatest increase in goodness-of-fit of the longitudinal multilevel models. Longitudinal analysis is a promising analytical strategy to identify reliable predictors of the clinical outcome of IBI.
© 2012 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.
All rights reserved.

PALABRAS CLAVE
Autismo;
predictor
Análisis aplicado de la conducta;
Cuasi-experimento (serie temporal interrumpida con un grupo)

Resumen La predicción de resultados de tratamiento es un componente importante de la planificación clínica. No obstante, no se han hallado predictores fiables de los efectos de la intervención conductual intensiva en personas con trastorno del espectro autista. La incorporación de análisis longitudinales a la investigación sobre resultados de tratamiento en este área puede contribuir a la identificación de predictores con valor clínico. En el presente estudio se evaluaron las habilidades verbales, cognitivas y de la vida diaria (Early Learning Accomplishment Profile y Learning Accomplishment Profile-Diagnostic, 3ª ed.) de 24 niños con trastorno del espectro autista en un programa de intervención conductual intensiva. Las evaluaciones se realizaron cada seis meses y durante un periodo medio de intervención de dos años. Mediante

*Corresponding author at: University of Manitoba, Psychology Department, P518 Duff Roblin Bldg., 190 Dysart Road, MB R3T Winnipeg, Manitoba, Canada.
E-mail address: javier.virues@ad.umanitoba.ca (J. Virues-Ortega).

1697-2600/$ - see front matter © 2012 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L. All rights reserved.

análisis multinivel se examinaron posibles predictores longitudinales incluyendo sexo, edad, intensidad y duración de la intervención, tiempo total de intervención y nivel de funcionamiento previo a la intervención. Los resultados indicaron que el tiempo total de intervención, el funcionamiento previo y la edad causaban los mayores incrementos en bondad de ajuste de los modelos longitudinales. El análisis longitudinal es una estrategia analítica prometedora en la identificación de predictores fiables de la efectividad de la intervención conductual intensiva.
© 2012 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L.
Todos los derechos reservados.

Autism spectrum disorder (ASD) is a pervasive developmental disorder that affects 1 to 2.5% of children (Baio, 2012). A number of comprehensive psychosocial interventions for people with ASD have been developed for which preliminary evidence exists. These include the Early Start Denver model (ESDM, Dawson et al., 2010), the Treatment and Education of Autistic and Related Communication Handicapped Children (TEACCH, Welterlin, Turner-Brown, Harris, Mesibov, & Delmolino, 2012), and intensive behavioral intervention based on the UCLA Young Autism Project model and applied behavior analysis (IBI, Lovaas, 1987). Although there is no single approach to treatment for all individuals with ASD, IBI based on applied behavior analysis is among the few approaches to treatment that have been tested extensively using clinical trial methodology (Rogers & Vismara, 2008; Virués-Ortega, 2010; Wetherby & Woods, 2006).

Applied behavior analysis is devoted to the experimental study of socially significant behavior as a function of environmental and social variables, and is the branch of experimental psychology that supports the conceptual framework of IBI (Luiselli, Russo, Christian, & Wilczynski, 2008). IBI is a comprehensive and evidence-based approach to the systematic teaching of behavioral, verbal, cognitive, and social repertoires to individuals diagnosed with ASD (Howlin, Magiati, & Charman, 2009). Treatment typically involves over 20 weekly hours of one-to-one teaching incorporating multiple learning trails and specific programs for targeted behavioral goals. Teachers program hundreds of learning trials per day featuring discrimination training, prompting, generalization, and other reinforcement-based procedures known to facilitate the acquisition of new skills in individuals with and without disabilities (Miltenberger, 2011). The IBI curriculum integrates complex sequences of programs from basic attending or vocalizing skills, up to complex verbal, social, and problem-solving skills (Lovaas, 2002).

Over 20 independent trials have been conducted which jointly suggest that IBI has moderate to large effects on daily living skills, cognitive functioning, language, and social behavior (Foxx, 2008; Remington et al., 2007; Virués-Ortega, 2010). The field of IBI has shown a considerable growth as suggested by the increasing number of service providers and certified professionals (Shook & Johnston, 2011).

Parents of children undergoing IBI and other evidence-based interventions frequently want to know whether their child will be able to attend school without special support, what areas of behavioral functioning - whether motor, social or cognitive - are likely to improve as a consequence of treatment, and what intervention intensity and duration may be optimal for their child. Until recently, outcome research had been of little assistance to respond to these and other questions pertaining to the longitudinal progression of children undergoing treatment.

While the evidence available strongly suggests that some individuals benefit significantly from IBI and other approaches to treatment, participant and intervention characteristics associated with greater intervention effects are not well understood. The wider literature of treatment outcomes in ASD has examined a range of mediating and moderating factors that could, potentially, be established as clinically valuable predictors. These include pre-intervention IQ, treatment duration and intensity, family characteristics, age at intervention onset, social initiation skills, and structural dismorphologies of the central nervous system. The scant literature available on these factors have been reviewed by Rogers and Vismara (2008) who concluded that "The current intervention research focus on main effects models provides little information about who does well in which treatments and why" (pp. 28-29).

Age, pre-intervention functioning, and intervention intensity have been examined in the narrower literature of IBI outcome predictors. Studies that have examined the role of age at the onset of IBI have shown that the earlier the intervention, the greater the intervention effect. For instance, Granpeesheh, Dixon, Tarbox, Kaplan, and Wilke (2009) found that children below seven years at treatment onset mastered more behavioral objectives every month than children who started IBI intervention above that age.

The studies that have examined pre-intervention functioning as a predictor of treatment outcome have not always been consistent in their findings. Perry et al. (2008) examined progress of children with ASD that received IBI services by comparing standardized assessments at the beginning and end of the service. Children were classified as having either higher, intermediate, or lower functioning at intake based on their Vineland Adaptive Behavior Composite score. The higher functioning group made substantial gains (~20 IQ increments) relative to the other two groups. By contrast, Ben-Itzchak, Lahat, Burgin, and Zachor (2008) reported that pre-intervention IQ (normal, borderline, low) did not predict the IQ gains after a year of IBI in a group of 81 young children with ASD and developmental disabilities.

More evidence has been accrued on the effects of intervention intensity. However, findings remain inconsistent. Taking IQ as a prototypical outcome (Table 1), Makrygianni and Reed (2010) in a correlational study did not find any effects of intensity – similar results were found by Sheinkopf and Siegel (1998). Virués-Ortega (2010)

Table 1 Effect of treatment intensity on IQ in intensive behavioral intervention outcome studies.

Study	Sample size[a]	Intensity range (h/week)	Analysis	Effect size
Makrygianni & Reed (2010)	86	15-30	Correlational (Pearson *r*)	.22
Sheinkopf & Siegel (1998)	11	21-32	Correlational (Pearson *r*)	−.06
Virués-Ortega (2010)	340	12-45	Meta-regression	.01

Note. Effects reported as Cohen d effect sizes. [a] Sample size of the intervention group.
*All effect sizes were non-significant, $p > .05$.

reported no effects of intensity on IQ in a pooled analysis of 19 experimental IBI studies. Finally, Reed, Osborne, and Corness (2007) established a moderate effect of intensity in a small trial on IBI using the Psychoeducational Profile as outcome. In summary, treatment intensity has not been established as a consistent predictor of IBI intervention effects.

Longitudinal modeling of intervention outcomes may help to establish intervention predictors more firmly than traditional pre-post assessments. Longitudinal analyses are able to fit the mathematical functions followed by outcome trajectories of individual clients over a period of time. By doing so longitudinal analysis maximizes the statistical power of regression models aiming at meaningful outcome predictors. For instance, if IBI effects were to follow a non-linear progression, rather than a linear trajectory, it may be possible to establish the role of a particular predictor more accurately through longitudinal multi-level analyses suited to specific non-linear mathematical functions (Singer & Willett, 2003). Furthermore, predictors identified based on time-series spanning the treatment duration, as opposed to pre-post assessments, may strengthen the clinical utility of the predictor. For example, pre-intervention functioning could be a strong predictor of treatment outcomes during the first year of treatment, but not during the second.

IBI operates through a package of systematic teaching strategies which are expected to provide the individual with an increasing set of cognitive and behavioral resources that will in turn offset, to various extents, the behavioral excesses and deficits that are characteristic of ASD and other developmental disabilities. Being a training-based and goal-directed approach to intervention, IBI may lead to some degree of behavioral gains for as long as the intervention is in place. Longitudinal analysis of IBI may help to identify distinct treatment gain itineraries across subjects and tie those to specific predictors. For instance, it may be possible that individuals starting at a higher pre-intervention level of functioning benefit more from IBI but reach an asymptote (ceiling) sooner than individuals that start at a lower level of functioning. The longitudinal predictors of IBI effects shall be greatly informative, albeit, they have been rarely explored in the literature. There are several longitudinal analyses that feature patterns of change in individuals with ASD (Dietz, Swinkels, Buitelaar, van Daalen & van Engeland, 2007; Jonsdottir et al., 2007; Magiati, Moss, Charman, & Howlin, 2011). Nonetheless, these analyses are constrained by the number of longitudinal assessments (three or less); the number of treatment outcomes (e.g., Dietz et al. only reported IQ); and the data analysis strategy (e.g., no multilevel analyses).

This article describes growth patterns of motor, cognitive, verbal, daily-living, and social skills in a sample of children with ASD admitted into a home-based IBI program managed by trained behavior analysts and delivering 20 to 40 weekly hours of intervention. We used the children's performance in standardized assessments conducted periodically to longitudinally create curves charting the rates and asymptotes of various behavioral repertoires. Subsequent analyses were conducted to test the impact of several personal and intervention-related predictors on the longitudinal growth of IBI outcomes. The present analysis may help to enhance the prognostic information available to families and clinicians by determining the extent to which specific client- and treatment-related variables more closely predict treatment outcome over the duration of the intervention.

Method

Participants

Twenty-four children diagnosed with ASD (Age: *Mean* = 50.05 months, *SD* = 28.3; Gender: 21 boys and 3 girls) admitted to the IBI program of *Fundación Planeta Imaginario* (Barcelona, Spain) participated in the study. An *a priori* power analysis indicated that a total sample size of 15 was required to detect large effects (Cohen effect size = 1). Therefore, our sample would suffice to identify moderate to large effect sizes. *A priori* power analysis assumptions were based on the pooled effect size of 20 trials on IBI using IQ reported by Virués-Ortega (2010) (Pooled effect size = 1.19). Participants were recruited consecutively and were not excluded based on their age or pre-intervention functioning at the time of referral. All participants received a diagnosis of ASD from an external medical consultant based on the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, 4th edition text revised. Diagnosis was supported by standardized assessments of autism including either the Autism Diagnostic Interview-Revised (ADI-R) or the Autism Diagnostic Observation Schedule-Generic (ADOS-G) (Le Couteur, Haden, Hammal, & McConachie, 2008). Further personal characteristics are presented in Table 2.

Table 2 Characteristics of the study sample.

	Pre-test (*N*=24)	Post-test (*N*= 24)
Age in months, *M*±*SD*	51.91±27.31	69.46±27.26
Gender (male:female)	23:1	
Ethnicity (% Caucasian)	100%	
Social class,[a] % high	100%	
IQ,[b] *M*±*SD*	74.50±13.98	91.50±16.86
Skills mastered in selected areas,[c] M±*SD*		
Attending (max. 19)	13.04±4.34	19.16±3.05
Imitation (max. 27)	7.84±8.41	19.92±7.24
Matching (max. 13)	6.02±7.48	13.08±6.08
Basic labeling (max. 13)	12.44±5.33	31.21±19.88
Independent play (max. 15)	3.76±4.76	11.72±6.00
Interaction with peers/adults (max. 25)	2.28±3.82	11.60±9.06

Note. [a]Estimated by parental education and professional background. [b]Wechsler Preschool and Primary Scale of Intelligence, 3rd ed.; Bailey Scales of Infant Development, and Merrill-Palmer Scales of Mental Tests. [c]Number of skills mastered by area (Lovaas Institute Midwest, 2010).
M = mean; SD = stardard deviation.

Instruments

Fine and gross motor, cognitive, language, self-care and social skills were assessed by means of the Early Learning Accomplishment Profile (E-LAP; Glover, Priminger, & Sanford, 1988; Peisner-Feinberg & Hardin, 2001) and the Learning Accomplishment Profile-Diagnostic, 3rd edition, (LAP-D; Hardin, Peisner-Feinberg, & Weeks, 2005). The E-LAP and LAP-D scores are developmental age values expressed in months. The score range is 0 to 36 for the E-LAP and 36 to 72 for the LAP-D. If a participant achieved the upper limit of the score range of E-LAP, the assessment would be repeated with the LAP-D, which would then continue to be used as the means of standardized assessment every 6-month period until treatment was discontinued. In order to control for potential ceiling effects in our data, if a participant reached the LAP-D ceiling, assessment could be repeated one additional time to inform maintenance (provided that the individual would continue to receive services through the program for the next six-month period).

Both the E-LAP and the LAP-D have a high level of inter-rater reliability, internal consistency, and convergent validity with IQ (Fleming, 2000; Hardin et al., 2005; Long, Blackman, Farrell, Smolkin, & Conaway, 2005; Peisner-Feinberg & Hardin, 2001). The test-retest reliability of both instruments is reportedly excellent, ranging between .93 and .99 (Peisner-Feinberg & Hardin, 2001, Hardin et al., 2005). Practice effects were unlikely, as exposure to materials and tasks during the assessment was minimal (few trials); and prompting, reinforcement, and correction strategies were not present during the assessment. The Spanish version of the E-LAP and the LAP-D materials were used in the present study. The LAP-D was validated in a representative sample of Spanish-speaking children (Hardin et al., 2005). No Spanish validation of the E-LAP is currently available. Nonetheless, test scoring is performance-based - there are no standard scores.

Both instruments have been used frequently as standardized assessments in intervention studies with individuals with ASD (e.g., Ganz, Simpson & Corbin-Newsome, 2008). Moreover, the construct validity of E-LAP and LAP-D is supported by items screening all diagnostic areas of ASD (e.g., "initiates on play activities," "responds correctly when asked to show a toy," "inflexible and rigid in behavior"), items informing non-pathognomonic clinical features of autism (e.g., motor functioning), and items covering developmentally relevant skills (e.g., matching skills). In summary, the E-LAP and LAP-D were considered adequate for the present analysis due to their likely resilience to practice effects; excellent stability; excellent convergent validity with intellectual assessment measures; and relevance to the clinical, adaptive, and behavioral features of ASD.

Procedure

Participants were admitted consecutively to an IBI program within the period May 2006 through January 2011. This program was an official international replication site of the UCLA Young Autism Project model and affiliated with the Lovaas Institute (2011). At the onset of intervention, participants received an average of 31.87 weekly hours (*SD* = 10.11, range 15 -47.30) of home-based systematic teaching following the UCLA young autism model of service delivery and curriculum (Lovaas, 2002). Average treatment duration was 21.87 months (*SD* = 14.38, range 5.33-58.57). In keeping with all IBI bonafide programs, in addition to the hours of formal intervention, incidental teaching and practice goals were operating during most waking hours (parents and caregivers acted as active co-therapists). One-to-one teaching was delivered by trained tutors that were supervised on a weekly basis by licensed psychologists

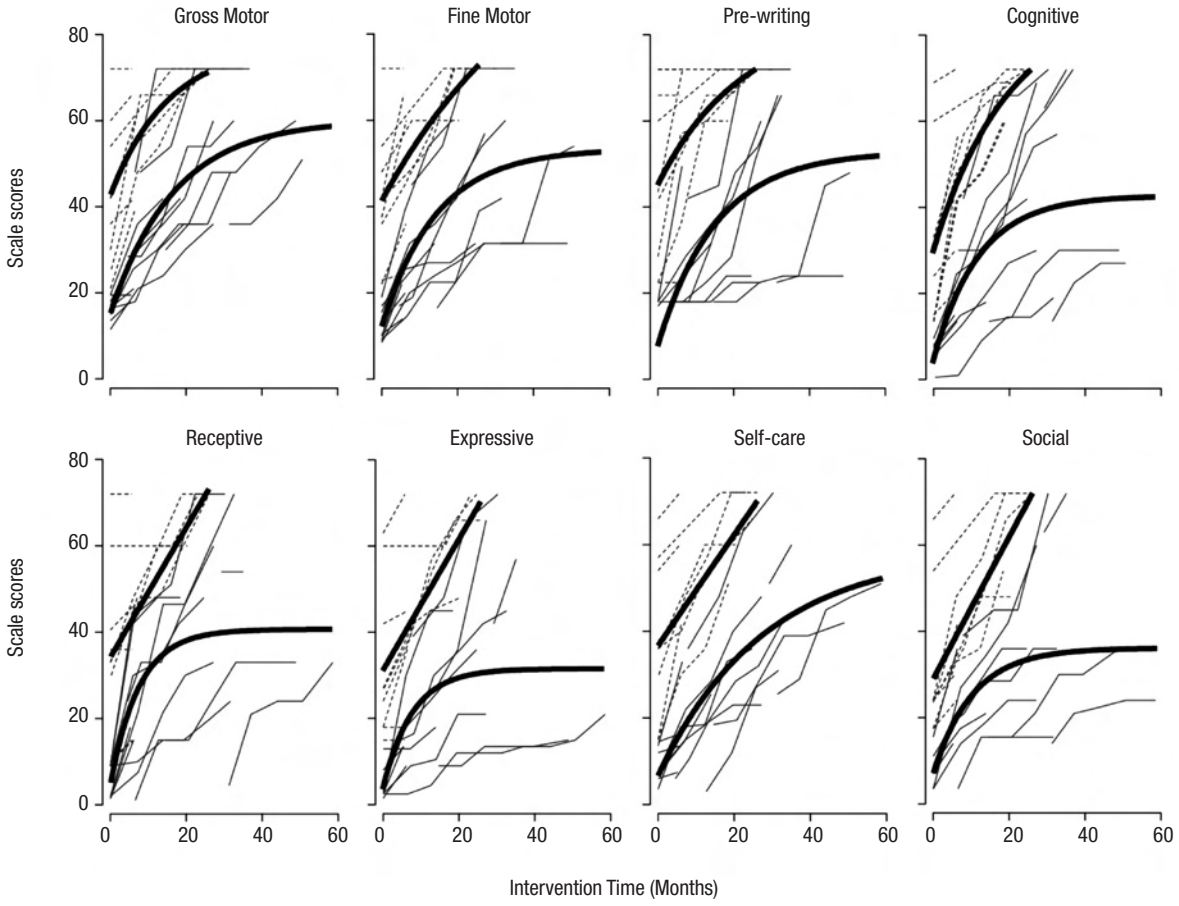


Figure 1 Trajectories of Early Learning Accomplishment Profile and Learning Accomplishment Profile-Diagnostic scores over time. Fitted exponential negative curves (solid black line) were obtained for individuals above (dotted grey lines) and below (solid grey lines) the median of pre-intervention functioning at baseline in each domain.

with a background in behavior analysis. Parents received weekly or bi-weekly progress updates, and supervision and specific routines that required their involvement in order to ensure the consistency of the interventions across contexts and caregivers. Intervention was individualized and comprehensive; and targeted motor, behavioral, daily-living, verbal, cognitive, and social skills. Goals were informed by a standardized curriculum composed of over 850 skills organized in 45 broad clinical areas (e.g., reading, self-control skills). These goals are informed by developmental sequences of typically developing children (Luiselli et al., 2008) and include skills that are instrumental for the acquisition of more complex repertoires (e.g., matching skills, imitation). Teaching sessions were delivered via one-to-one teaching with gradual transition to group activities and natural contexts. Transition to natural social contexts was emphasized after mastery in one-to-one teaching format. Decision-making in terms of hour allocation and treatment discontinuation weighted a number of factors including availability of school support, progress achieved, family priorities, and treatment costs. Typically, individuals that showed a persistent asymptote in their learning achievements or that became independent at school were assigned a reduced number of hours in preparation of service discontinuation (for details on the IBI curriculum see Lovaas, 2002). The current program was in line with the guidelines for responsible conduct published by the Behavior Analyst Certification Board (2010).

All participants underwent standardized assessments with the E-LAP or the LAP-D prior to the intervention and approximately every six months into the program (average data points per participant 3.8, range 2-6). The selection, administration, and correction of instruments followed the guidelines by Jurado and Pueyo (2012).The research assistants conducting the standardized assessments were not involved in the administration of treatment and were not familiar with the hypotheses of the study.

Data analysis

Figure 1 shows the individual growth trajectories of participants for the eight E-LAP and LAP-D outcomes. Visual inspection of the data plots over time suggests that trajectories accelerated away from the start point shortly after the intervention commenced while progression decelerated as the individual approached a personal or scale ceiling. Therefore, individual trajectories did not follow a linear progression but rather an exponential negative growth. Exponential negative trajectories are composed formally of a negatively accelerated curve,

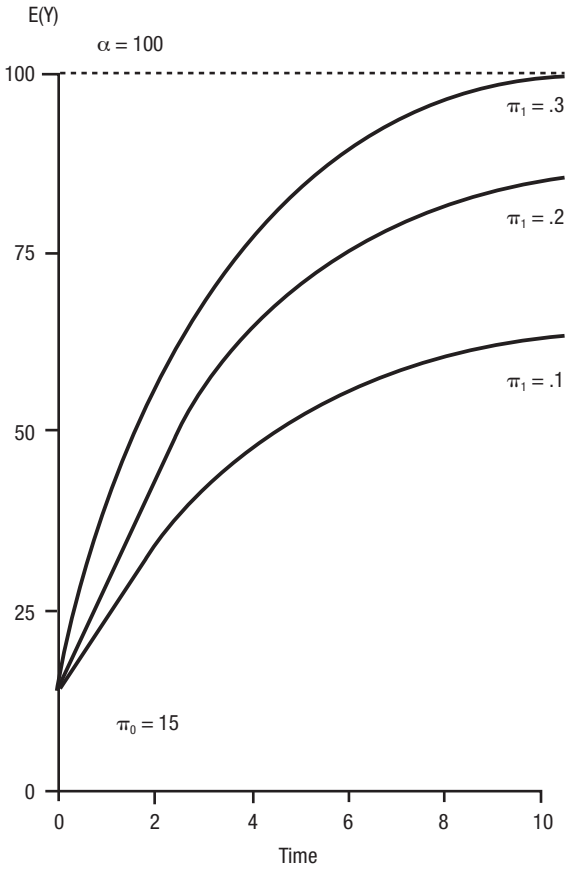


Figure 2 Exponential patterns of change based on different parameter values.

ending in an upper asymptote. According to the formal attributes of the data we selected a multilevel regression model based on the following exponential negative function:

$$Y_{ij} = \alpha_i - (\alpha_i - \pi_{0i})\, e^{-\pi\, TIMEij}$$

Where α_i represents the upper asymptote, π_{0i} represents the lower end of the trajectory, and π_{1i} represents the slope of the curve. Figure 2 illustrates different exponential negative patterns of change over time for various parameter values.

Multilevel models provide two distinct levels of analysis: level-1 and level-2. The structural parts of the level-1 submodel contain two level-1 parameters and one within-person variance component (ε_{ij}). The first parameter, known as *intercept* (π_{0i}), represents the initial status of an individual *i* in the population. The second parameter, known as *slope* (π_{1i}), represents the rate of change for the individual *i* in the population by unit of time. Therefore, level-1 establishes individual change overtime. By contrast, the parameters at level-2 do not represent individual variation, but average level of the outcome in the population. Specifically, the parameters at level-2 represent the average outcome level in the population corresponding to the intercept and slope values at level-1. At level-2, the pattern of change is not examined in terms of time, as is the case at level-1, but rather, in terms of a predictor. In summary, there are four parameters at level-2: γ_{00} is the population average of level-1 intercept with level-2 predictor value of 0, γ_{01} is the population average difference in level-1 intercept for a 1-unit variation in the predictor, γ_{10} is the population average of the level-1 slope when the predictor equals 0, and finally, γ_{11} is the population average difference when the predictor equals 1. γ_{00} and γ_{10} are baseline parameters while γ_{01} and γ_{11} estimate the association of the predictor with the initial status and the rate of change of the longitudinal progression, respectively. The model also provides a residual variance value for the intercept (σ_0^2), the slope (σ_1^2), and the covariance among these two (σ_{01}). For multilevel models incorporating two predictors we will also report γ_{12} and γ_{12}, which represents the population's average variation in the outcome level for a one-unit increment in the predictors 1 and 2 (level-1), respectively (for more details in multilevel analysis refer to Singer & Willet, 2003). The estimation of the predictor coefficients at level-2 is presented formally below:

$$\pi_{0i} = \Upsilon_{00} + \Upsilon_{01}\,(PREDICTOR_i - PREDICTOR) + \xi_{0i}$$
$$\pi_{0i} = \Upsilon_{10} + \Upsilon_{11}\,(PREDICTOR_i - PREDICTOR) + \xi_{1i}$$

According to this model, individual growth parameters (π_{0i}, π_{1i}) across children will be a function of population average values (γ_{00}, γ_{10}), and population variance components (ξ_{0i}, ξ_{1i}) represented by residual variances (σ_0^2, σ_1^2) and covariance (σ_{01}).

We estimated a series of multilevel models using different sets of predictors in order to select models that would maximize goodness-of-fit for a given outcome when compared with an unconditional baseline model (model with no predictors). This was accomplished in two sequential multilevel analyses. In the first sets of models we examined the impact of time-based predictors (intervention duration in weeks, total hours of intervention - weekly hours of interventions multiplied by weeks of intervention - and age in months). We would then select the model incorporating the single time-based predictor with best goodness-of-fit for each of the eight outcomes under analysis. Subsequently, we calculated a new set of two-predictor models incorporating the predictor previously selected and a specific personal factor that, when added, resulted in further increases in goodness-of-fit. The personal factors examined for each of the eight outcomes were age (if not selected in the preceding step), gender, and pre-intervention functioning. Two levels of pre-intervention functioning were established using the median value at baseline as cut-off point. The rationale for selecting these predictors is twofold: a) they are all common individual/treatment characteristics readily accessible to the clinician, and b) they have been examined in previous IBI studies although not in the context of a longitudinal analysis. Longitudinal predictors that changed overtime (intervention duration, total intervention duration, age) were re-calculated each time an individual was assessed.

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were computed as goodness-of-fit parameters for all one- and two-predictor models. Lower AIC and BIC values are indicative of better fitting. The best

Table 3 Goodness-of-fit parameters of all one- and two-predictor multilevel models of change for Early Learning Accomplishment Profile and Learning Accomplishment Profile-Diagnostic scores.

	Goodness-of-fit (AIC, BIC)							
	GMF	**FMF**	**PWR**	**COG**	**RLG**	**ELG**	**SFC**	**SBH**
Unconditional model	761.56	774.54	761.31	774.29	809.53	822.50	782.89	795.87
	800.47	813.44	761.91	774.88	755.07	768.05	747.29	760.26
One-predictor models								
Intervention duration	715.86	721.17	779.29	746.27	763.16	732.19	706.67	706.42
	731.43	736.70	794.86	761.84	778.73	747.77	722.24	721.99
Total intervention time	693.33	691.12	744.67	720.03	737.42	712.48	677.15	688.53
	708.59	706.38	759.93	735.29	752.68	727.74	692.41	703.79
Age	716.29	731.43	776.85	761.27	769.47	740.37	708.01	721.45
	731.86	747	792.42	776.84	785.03	755.94	723.58	737.02
Two-predictor models								
Age	664.01	673.20	727.92	710.50	724.67	702.71	651.35	676.02
	681.81	691	745.72	728.30	742.47	720.52	669.16	693.82
Gender	691.93	690.19	745.08	719.15	735.78	711.83	675.18	687.74
	709.73	707.99	762.88	736.95	753.59	729.63	692.99	705.55
Pre-intervention level	675.54	661.39	715.56	705.10	726.15	698.89	661.74	676.27
	693.35	679.20	733.36	722.90	743.96	716.70	679.55	694.07

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; COG = Cognitive; ELG = Expressive language; FMF = Fine motor function; GMF = Gross motor function; PWR = Prewriting; RLG = Receptive language; SBH = Social behavior; SFC = Self-care. Intercept constant; slope is established as the intervention duration in months. Level-2 and level-3 best fitting models by outcome are highlighted.

fitting two-predictor model was selected for each outcome and was fully reported. All analyses were conducted with STATA version 11 (STATA Corporation, College Station, TX) and its GLAMM program for multi-level analysis. A .05 level of significance was used throughout. Results have been reported according to the guidelines by Hartley (2012).

By comparing the goodness-of-fit of one- and two-predictor models with an unconditional model, we aimed to establish which factors would better explain the longitudinal variation in our data. This analysis will help to determine prominent trajectories of intervention outcomes based on specific predictors. This strategy also serves the purpose of suggesting causality in the absence of a control group, similar to the way in which dose-response relations inform causation (see a discussion relevant to this point in Arjas and Parner, 2004). Namely, the causation inference would be supported if intervention intensity (e.g., total intervention hours at each time of assessment) is indeed superior in its ability to increase the fit of the model relative to an arbitrary time-dependent predictor (individuals' age).

Results

The examination of the goodness-of-fit parameters of multilevel regression models showed that one-predictor and two-predictor models had a superior fit than unconditional models for every domain of the E-LAP and the LAP-D. AIC and BIC goodness-of-fit parameters of all models are reported in Table 3. Total intervention time (hours per week multiplied by weeks of intervention) was the single predictor with the highest favorable impact on goodness-of-fit for all E-LAP and LAP-D outcomes. Other time-based predictors including individuals' age and intervention duration in months had a positive impact in the model's fit, but did so to a lesser extent than total intervention time in all eight outcomes.

Further improvements in goodness-of-fit were achieved in two-predictor models. Keeping total intervention time as the first factor, we examined the fit of regression models incorporating age, gender, or pre-intervention level as a second predictor. Age was the second most efficient predictor in terms of improving fit of the regression models for gross motor function, receptive language, self-care, and social behavior; while pre-intervention level was the second most efficient predictor for regression models using fine motor function, prewriting, cognitive, and expressive language (Table 3). The regression models of domains assessing motor, daily living, and social skills (gross motor function, fine motor function, self-care and social behavior) achieved better fitting than regression models of language-related domains (prewriting, receptive language, expressive language, cognitive).

Table 4 presents the best fitting two-predictor multilevel model for each E-LAP and LAP-D outcome. Both predictors were statistically significant ($p < .001$) for every outcome. Rate of change attributable to total intervention time in hours (γ_{11}) ranged from .004 to .009 (outcome average increase by predictor unit). Coefficient magnitudes for age in months (γ_{12}) as a predictor ranged from .391 to .514. Finally, coefficients for the dichotomous variable pre-

Table 4 Multilevel models for Early Learning Accomplishment Profile and Learning Accomplishment Profile-Diagnostic scores change over the duration of intensive behavioral intervention.

	Goodness-of-fit (AIC, BIC)							
	GMF	**FMF**	**PWR**	**COG**	**RLG**	**ELG**	**SFC**	**SBH**
Fixed effects								
Intercept (γ_{00})	7.44	13.91**	10.97*	8.87	−3.00	5.46	-4.45	.52
Intervention, hours (γ_{11})	.00**	.01**	.01**	.01**	.01**	.01**	.01**	.01**
Age, months (γ_{12})	.51**	-	-	-	.51**	-	.57**	.39**
Pre-intervention level[a] (γ_{12})	-	28.43**	35.67**	24.27**	-	22.97**	-	-
Variance components								
Level-1: Within-person (σ_ε^2)	25.32	23.32	47.64	32.91	47.79	28.39	16.54	18.90
Level-2: Intercept (σ_0^2)	89.91	60.71	158.79	123.89	204.10	122.62	155.74	159.19
Level-2: Slope (σ_1^2)	.07	.19	.25	.33	4.37	.40	.10	.32
Level-2: Covariance (σ_{01})	.46	2.37	-4.16	2.76	.11	.98	2.81	3.92

Note. COG = Cognitive; ELG = Expressive language; FMF = Fine motor function; GMF = Gross motor function; PWR = Prewriting; RLG = Receptive language; SBH = Social behavior; SFC = Self-care. Goodness-of-fit parameters and domain abbreviations in Table 3. *p < .01; **p < .001. [a]Pre-intervention levels above and below the median at pre-test.

intervention level (γ_{12}) ranged from 22.971 to 35.669. Figure 1 portrays fitted curves based on an exponential negative growth of subsamples above and below the median value of pre-intervention level for each of the eight standardized outcomes.

Discussion

Multilevel regression analyses based on an exponential negative growth trajectory indicated that total intervention duration in hours was the single predictor with the highest contribution to the model fit for all outcomes when compared with unconditional models. This finding suggests that a subtle characteristic of the intervention – a combination of both treatment intensity (weekly hours) and treatment duration (total weeks of treatment) – optimizes the fitting of individual trajectories to a specific mathematical function for the duration of the intervention and across a range of standardized outcomes. Improvements in model fitting caused by duration alone did not improve goodness-of-fit to the extent achieved by total intervention time as a single predictor (Table 3). Therefore, our data suggest that both intensity and duration, as represented by total intervention time, remained important factors of intervention gains regardless of pre-intervention functioning or age. Finally, total intervention time remained significant (p < .001) in all final two-predictor multilevel models (Table 4). When used in one-predictor models, pre-intervention functioning was inferior to total intervention time in terms of improving goodness-of-fit for all outcomes.

We tested the impact of pre-intervention functioning in the goodness-of-fit of multilevel models incorporating two predictors. Including pre-intervention level as a second predictor, improved goodness-of-fit for all outcomes in the two-predictor models (Table 3). For four of the eight standardized outcomes examined (fine motor, pre-writing, cognitive, expressive language), pre-intervention level was the personal characteristic (above age and gender) that generated the greatest improvement in model fit. Pre-intervention level was a significant factor (p < .001) in the final two-predictor models for fine motor, pre-writing, cognitive and expressive language domains (Table 4). Interestingly, these outcomes involved more complex cognitive abilities relative to the remainder of E-LAP and LAP-D outcomes (e.g., fine vs. gross motor; expressive vs. receptive language; cognitive vs. self-care).

Our results suggest that individuals starting intervention at a lower level in a given outcome were more likely to follow an asymptotical growth as opposed to individuals that initiated treatment with a higher level of performance (cf. fitted curves on Fig. 1). The visual inspection of the individual longitudinal trajectories in our sample suggests that pre-intervention level is a plausible predictor of individuals' performance over the course of the intervention to the extent that a bimodal pattern seems obvious in most of the outcomes (e.g., Cognitive, Social). Bimodal trajectories in our dataset are consistent with the distinction between most and least positive responders to IBI discussed by Remington et al. (2007). The visual examination of individual trajectories on Figure 1 suggests that the pre-intervention median is an acceptable cut-off point as attested by the predictors significance and fit gains in models that incorporated this factor. A more sophisticated strategy to determine the cut-off point would have required asymmetrical assignment of participants above and below the cut-off points, which may have harmed statistical power and increase the potential for type II error. Therefore, future analyses would benefit from samples sizes larger than ours.

Learning processes have been found to accommodate well to exponential negative or logistic patterns of change (e.g., Hicklin, 1976). The possibility remains, however, that non-linear patterns of growth found in the present study may have been caused by measurement-dependent factors, like inadequate scaling assumptions or excessive ceiling

effects in the psychometric instrument used to establish treatment outcomes. These potential shortcomings, however, may have had little impact on the validity of the predictors, which is independent from the specific shape of the longitudinal growth.

The contributions of our study are primarily methodological and to a lesser extent practical. As discussed in our introduction, the literature on the effect of intensity and other predictors on the outcome of IBI have yielded inconsistent results. This inconsistency may be explained, at least to some extent, by non-linear variations of the predictor and the outcome overtime. Therefore, longitudinal studies may enhance our ability to examine outcome predictors with sufficient statistical power. Our results provide evidence in this direction being the first study to use this methodology in the context of IBI intervention.

In terms of the applied relevance of our findings, future longitudinal studies expanding the present analysis could eventually provide the basis for evidence-informed clinical decision-making. Namely, clinicians could combine various predictors available at the beginning of the intervention (e.g., pre-intervention functioning in an specific area, age, expected treatment intensity and duration) to estimate the progress of the client over the next years, which could in turn inform the decision-making of family, caregivers and health decision-makers in terms of treatment planning and resource allocation.

In summary, the present analysis helps to identify the general features of the longitudinal progression of children with autism undergoing IBI. Our results suggest that increased intervention time, lower age at intervention onset, and higher pre-intervention functioning might be associated with greater IBI outcomes for intervention programs of up to four years in duration. The present study provides the methodological basis for predictor identification in the longitudinal analysis of IBI.

References

Arjas, E., & Parner, J. (2004). Causal reasoning from longitudinal data. *Scandinavian Journal of Statistics, 31*, 171-187.

Baio, J. (2012). Prevalence of autism spectrum disorders: Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveillance Summaries, 61*(SS03), 1-19.

Behavior Analyst Certification Board (2010). *BACB Guidelines for responsible conduct for behavior analysts*. Available from: http://www.bacb.com/index.php?page=57 [retrieved 23 Jan 2013].

Ben-Itzchak, E., Lahat, E., Burgin, E., & Zachor, A. D. (2008). Cognitive, behavior and intervention in young children with autism. *Research in Developmental Disabilities, 29*, 447-458.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver model. *Pediatrics, 125,* 17-23.

Dietz, C., Swinkels, S. H., Buitelaar, J. K., van Daalen, E., & van Engeland, H. (2007). Stability and change of IQ scores in preschool children diagnosed with autistic spectrum disorder. *European Child & Adolescent Psychiatry, 16,* 405-410.

Fleming, J.A. (2000). *An examination of inter-rater reliability of the E-LAP*. Baltimore: Johns Hopkins University.

Foxx, R.M. (2008). Applied behavior analysis treatment of autism. *Child and Adolescent Psychiatric Clinics of North America, 17*, 821-834.

Ganz, J. B., Simpson, R. L., & Corbin-Newsome, J. (2008). The impact of the picture exchange communication system on requesting and speech development in preschoolers with autism spectrum disorders and similar characteristics. *Research in Autism Spectrum Disorders, 2*, 157-169.

Glover, M. E., Priminger, J. L. & Sanford, A. R. (1988). *Early learning accomplishments profile*. Winston-Salem, NC: Kaplan.

Granpeesheh, D., Dixon, D. R., Tarbox, J., Kaplan, A. M., & Wilke, A. E. (2009). The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 3*, 1014-1022.

Hardin, B. J., Peisner-Feinberg, E. S., & Weeks, S. W. (2005). *The Learning Accomplishment Profile-Diagnostic (LAP-D), third edition*. Lewisville, NC: Kaplan Early Learning.

Hartley, J. (2012). New ways of making academic articles easier to read. *International Journal of Clinical and Health Psychology, 12*, 143-160.

Hicklin, W. J. (1976). A model for mastery learning based on dynamic equilibrium theory. *Journal of Mathematical Psychology, 13*, 79-88.

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities,114*, 23-41.

Jónsdóttir, S. L., Saemundsen, E., Asmundsdóttir, G., Hjartardóttir, S., Asgeirsdóttir, B. B., Smáradóttir, H. H., Sigurdardóttir, S., & Smaári, J. (2007). Follow-up of children diagnosed with pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 37*, 1361-1374.

Jurado, M. A., & Pueyo, R. (2012). Doing and reporting a neuropsychological assessment. *International Journal of Clinical and Health Psychology, 12*, 123-141.

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders, 38*, 362-372.

Long, C. E., Blackman, J. A., Farrell, W. J., Smolkin, M. E. & Conaway, M. R. (2005). A comparison of developmental versus functional assessment in the rehabilitation of young children. *Pediatric Rehabilitation, 8*, 151-161.

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.

Lovaas, O. I. (2002). *Teaching developmentally disable children: The ME Book*. Austin, TX: Pro-ed.

Lovaas Institute (2011). *NIMH Replication Sites*. Available from: http://www.lovaas.com/contact.php [retrieved 1 Sep 2012].

Lovaas Institute Midwest (2010). *Matrix curriculum and clinical goal areas*. Minneapolis, MN: Author.

Luiselli, J. K., Russo, D. C., Christian, W. P., & Wilczynski, S. M. (2008). *Effective practices for children with autism*. New York: Oxford University Press.

Magiati, I., Moss, J., Charman, T., & Howlin, P. (2011). Patterns of change in children with autism spectrum disorders who received community based comprehensive interventions in their pre-school years: A seven year follow-up study. *Research in Autism Spectrum Disorders, 5*, 1016-1027.

Makrygianni, M. K., & Reed, P. (2010). Factors impacting on the outcomes of Greek intervention programmes for children with autistic spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 697-708.

Miltenberger, R. G. (2011). *Behavior modification: Principles and procedures* (5th ed.). Belmont, CA: Cengage.

Peisner-Feinberg, E. S., & Hardin, B. J. (2001). *The early learning accomplishment profile edition examiner's manual and technical report*. New York: Kaplan.

Perry, A., Cummings, A., Geier, J. D., Freeman, N. L., Hughes, S., LaRose, L., Managan, T., Reitzel, J. A., & Williams, J. (2008). Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 2*, 621-642.

Reed, P., Osborne, L. A., & Corness, M. (2007). Relative effectiveness of different home-based behavioral approaches to early teaching intervention. *Journal of Autism and Developmental Disorders, 37*, 1815-1821.

Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., Alsford, P., Lemaic, M., & Ward, N. (2007). Early intensive behavioral intervention. *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.

Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders, 28*, 15-23.

Shook, G. L., & Johnston, J. M. (2011). Training and professional certification. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 498-510). New York: Guilford Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood. *Clinical Psychology Review, 30*, 387-399.

Welterlin, A., Turner-Brown, L. M., Harris, S., Mesibov, G., & Delmolino, L. (2012). The home TEACCHing program for toddlers with autism. *Journal of Autism and Developmental Disorders, 42*, 1827-1835.

Wetherby, A. M., & Woods, J. J. (2006). Early social interaction project for children with autism spectrum disorders beginning in the second year of life: A preliminary study. *Topics in Early Childhood Special Education, 26*, 67-82.