Check for updates

Behavioral Interventions

WILEY

RESEARCH ARTICLE

Changes in Curriculum Assessment Scores Associated With Early Intensive Behavioral Intervention in a Routine Clinical Care Setting

Linda A. LeBlanc[1] | Brent A. Kaplan[2] | Kelsie Thompson[1] | Alec Sweatman[1] | Charna Mintz[1]

[1]Action Behavior Centers, Austin, Texas, USA | [2]codedbx, Lexington, Kentucky, USA

Correspondence: Linda A. LeBlanc (linda.leblanc@actionbehavior.com)

Received: 2 April 2026 | **Revised:** 24 August 2026 | **Accepted:** 28 August 2026

Keywords: autism | curriculum assessment | early intervention | VB-MAPP | Vineland

ABSTRACT

Early intensive behavioral intervention (EIBI) is designed to establish critical social, cognitive, language, and adaptive repertoires. The current study examines changes in scores and growth curves on the most commonly used curriculum assessment, the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP), in a routine clinical care setting for autistic children in EIBI services. It is also the first study to examine how correlations between VB-MAPP scores and the Vineland Adaptive Behavior Scale–Third Edition (VABS-III) (i.e., a norm-referenced measure) throughout treatment. The mean VB-MAPP change score was 69.6 points in about 1.5 years of treatment, with the greatest change occurring between the first and second assessments (i.e., initially steep growth curves that lessen in slope over time). The VABS-III communication domain, adaptive behavior composite, and socialization domain had moderate to strong correlations with VB-MAPP scores and those correlations became stronger throughout the treatment process. Age was a predictor of initial VB-MAPP scores (i.e., lower age, lower score), supporting the importance of early identification and access to treatment. Children who exited services prematurely made progress in treatment, but more slowly than children who remained in services until a recommended discharge.

1 | Introduction

The language and social-cognition challenges experienced as part of autism spectrum disorder (ASD) can exert pervasive influence on overall development, cognition, and adaptive functioning (American Psychiatric Association 2022; Volkmar et al. 2014). Advances in awareness and diagnostic and screening tools have resulted in increased identification of ASD (i.e., 1 in 31; Shaw et al. 2025) and increased access to early intervention services, which is associated with better long-term outcomes for children with ASD and their families (Hyman et al. 2020).

Behavior analysis refers to the scientific study of principles of learning and behavior, focusing on how environmental factors influence actions (Behavior Analyst Certification Board, n.d.; Cooper et al. 2020). The applied branch of this science, applied behavior analysis (ABA), uses the principles of the science to develop therapeutic techniques (e.g., positive reinforcement) and treatment approaches for changing behavior and teaching skills. Treatment approaches based on ABA have been empirically shown to be effective in a wide variety of areas, including ASD (Behavior Analyst Certification Board, n.d.; Eldevik et al. 2009; Reichow and Wolery 2009; Reichow et al. 2018; Wergeland et al. 2022).

The term early intensive behavioral intervention (EIBI) refers to a specific model of ABA treatment where services occur intensively (e.g., 30–40 h per week) and at an early age (e.g., ages 2–6 years) for multiple years (Council of Autism Service

© 2026 John Wiley & Sons Ltd. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

Providers 2024; Virués-Ortega 2010). This treatment model focuses on all aspects of functioning and is designed to establish critical language, social, cognitive, and adaptive repertoires that neurotypical children rapidly develop beginning before age 2 (MacDonald et al. 2014). Significant improvements on the primary outcome metrics examined in these studies (e.g., intellectual or developmental quotient scores, adaptive behavior standard scores, standardized language scores; Ridout and Eldevik 2024) matter because the purpose of EIBI-services is to establish learning repertoires and alter developmental trajectory to catch children up to their same age peers (Klintwall et al. 2015; MacDonald et al. 2014). The scores themselves are not the important outcome; however, greater change in scores and higher scores at discharge are associated with less restrictive educational placements (e.g., Waters et al. 2020; LeBlanc et al. 2025) and greater learning and adaptive abilities (e.g., Howard et al. 2014; H. Cohen et al. 2006), which can have cascading positive effects on the child and family across their lifespan of continued development.

The effectiveness of EIBI for each of these areas has been well established across multiple decades (H. Cohen et al. 2006; Eldevik et al. 2026; Frazier et al. 2021; Howard et al. 2014; Reichow 2012) with evidence that an earlier age at the outset of services (e.g., MacDonald et al. 2014), baseline level of functioning (e.g., Ben-Itzchak et al. 2014; Smith et al. 2015) and greater treatment intensity (Eldevik et al. 2026) produce better outcomes. Outcomes on adaptive functioning are particularly important because adaptive functioning scores have been found to be significantly lower than expected for corresponding IQ scores for individuals with ASD (Alvares et al. 2020) but are more highly associated with adult outcomes than measures of cognitive functioning (Farley et al. 2009). In the absence of treatment, adaptive functioning scores often decrease, and the cognitive/adaptive functioning gap increases with increasing age increases (Kanne et al. 2011; Casula et al. 2024; Szatmari et al. 2015). However, with EIBI adaptive functioning scores can improve, and the largest changes are often seen in the 1st year of treatment (Howard et al. 2014) and at high intensity of treatment services (Eldevik et al. 2026).

Although most of the early controlled studies on EIBI were conducted in research settings, recent studies have demonstrated positive effects in routine clinical care settings (LeBlanc et al. 2025; Waters et al. 2020). Studies in routine clinical care settings attempt to evaluate the effectiveness, rather than efficacy, of an evidence-based intervention under the "real world" conditions encountered by the average person with their local provider rather than in a research setting (Wergeland et al. 2022). Research settings and routine clinical care often differ in terms of the characteristics of the patients (e.g., more heterogeneous) and treatments (e.g., patient attrition resulting in shorter duration of treatment), and other factors (e.g., therapists, assessments, supervision, and training) which can impact treatment results (Weisz et al. 2013). Since the majority of children with ASD receive their treatment services in routine clinical care sites rather than research settings, additional investigations of the outcomes achieved in these settings are needed to determine how these results compare with published outcomes from research settings.

1.1 | Verbal Behavior Assessments

Skinner's analysis of language, detailed in his book *Verbal Behavior* (1957), has become a commonly used framework for assessment and intervention programming in EIBI. This framework focuses on a functional assessment of the antecedents and consequences that evoke and maintain various aspects of language, with each unique set of antecedents and consequences referred to as a verbal operant (e.g., mand, tact, echoic, intraverbal, listener). Multiple behavior analytic assessment protocols are grounded in Skinner's verbal operant framework (Belisle et al. 2022; Dixon et al. 2015) and are used to guide behavior analytic language and social programming for young children with autism. Padilla (2020) surveyed the assessment practices of behavior analysts and found that the most commonly used assessment for identifying language deficits and needs to guide programming is the Verbal Behavior Milestones Assessment and Placement and Program (VB-MAPP; Sundberg 2014) with 76% of respondents reporting use of the VB-MAPP.

The VB-MAPP Milestones Assessment is used to evaluate 170 key developmental social and language milestones that typically emerge during the first 4 years of life. The first level samples prerequisite skills that typically develop in the first 18 months (e.g., requesting/mand), whereas the second level expands to skills developed between 18 and 30 months (e.g., advanced mands, intraverbals, tacts). Finally, the third level evaluates skills typically acquired between 30 and 48 months, including complex mands and tacts, visual-perceptual skills, and enhanced social interactions.

The VB-MAPP is a curriculum assessment (i.e., an assessment of the skills that can be targeted in treatment) rather than a standardized (i.e., norm-referenced, criterion-referenced) test. Although the administrator's manual and some researchers have described the VB-MAPP as a criterion-referenced assessment (Lotfizadeh et al. 2025; Padilla et al. 2023; Sundberg 2014). The manual does not describe published or unpublished analyses comparing VB-MAPP scores to an externally established set of interpretive criteria, as is required for criterion-referenced assessments. Instead, the assessment focuses on the skills that can be targeted in an early intervention curriculum rather than providing comparisons to a normative sample or to a specific cut-off score. Though not a standardized assessment, VB-MAPP Milestones scores indicate progress in the skills targeted in treatment and are typically reported to funders to document patient progress and need for services. Thus, it is important to understand whether change scores reflecting progress in the treatment program relate to improvements on standardized outcome measures that have been used to document developmental catch up in research evaluations of EIBI.

The reliability and validity of the VB-MAPP are currently being established. Montallana et al. (2019) reported intraclass correlation coefficients for the VB-MAPP, indicating moderate to good reliability for the overall Milestones assessment (38.1%) and for the vast majority of the individual domains (i.e., 56.3% of domains had moderate reliability, 31.3% good reliability, 6.3% had excellent reliability). Dixon et al. (2015) compared the VB-

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MAPP Milestones to another behavioral language and learning assessment, the Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System (Dixon 2014), and found significant correlations between the total scores of each measure, suggesting convergent validity. Padilla and Akers (2021) examined the content validity of the VB-MAPP using a panel of 13 subject matter experts to rate the (a) domain relevance, (b) developmental age appropriateness, and (c) appropriateness of the measurement method. The Milestones Assessment (i.e., all items in all domains in all three developmental levels) had a content validity estimate of 0.32, indicating that the majority of SMEs rated the items as "essential". Similar content validity estimates were found for each individual level: 0.35 (level 1), 0.30 (level 2), and 0.30 (level 3). In addition, they found moderate to strong evidence for age appropriateness and the method of measurement appropriateness for the vast majority of Milestones domains.

Lotfizadeh et al. (2025) conducted an archival analysis comparing VB-MAPP scores and Vineland Adaptive Behavior Scales–Second Edition (VABS-II) Comprehensive Interview Form (Sparrow et al. 2005) scores at a single assessment interval for 235 children with autism aged 3–9 years in a routine clinical care setting. The VB-MAPP Milestones assessment correlated moderately or strongly with raw scores on the VABS-II Communication ($r = 0.79$), Socialization ($r = 0.55$), and Daily Living Skills ($r = 0.55$) subdomains, as well as the overall raw score ($r = 0.72$) and adaptive behavior composite standard score (ABC; $r = 0.44$), which was interpreted as evidence for convergent validity of the VB-MAPP. In contrast, the Milestones assessment was weakly correlated with the VABS-II Motor Skills domain raw score ($r = 0.40$), which was interpreted as evidence of divergent validity. In summary, preliminary evidence exists for the reliability and validity of the Milestones assessment of the VB-MAPP.

Researchers have begun to examine changes in VB-MAPP scores associated with ABA treatment at various ages and intensities. Vietze and Lax (2020) studied the impact of low intensity ABA (i.e., up to 10 h a week) for over 100 toddlers with ASD in a preschool setting. These children ranged in age from 20 to 40 months ($M = 28.68$ months) at the start of services and remained in services for an average of about 10 months. All children improved on VB-MAPP scores, with a statistically significant increase from a mean pretest score of 28.31 to a mean posttest score of 73.14 (i.e., a 44.8-point difference). In addition, they found that younger children experienced bigger gains than older children, and children who received fewer total hours of treatment benefited more if they were younger at the start of services.

Lotfizadeh et al. (2020) conducted an archival analysis and found that low intensity in-home ABA services produced greater gains on the VB-MAPP after 2 years ($M = 39.7$; SD = 16.4) than a comparison group who received eclectic treatment with fewer ABA treatment hours ($M = 31.1$; SD = 16.9). Since the treatment group began with a lower average VB-MAPP score, intake scores were statistically controlled using an ANCOVA, resulting in a statistically significant effect of intervention type and intensity on VB-MAPP scores. They excluded participants who had higher VB-MAPP scores at intake (i.e., 62% of all the VB-MAPP Milestones at intake) and only assessed at the 2-year mark, which limits our understanding of improvement for children with higher scores at intake and of how scores changed throughout the treatment process (i.e., growth curves).

Although growth curve analyses have not been conducted with the VB-MAPP, Myers et al. (2022) examined growth curves on the Assessment of Basic Language and Living Skills–Revised (ABLLS-R; Partington 2010) for a sample of 352 children with ASD. The ABLLS-R was administered at intake and approximately every 4 months throughout treatment for up to 10 administrations. The initial (i.e., time 1) ABLLS-R average composite score had doubled 8 months later (i.e., time 3), and the rate of increase accelerated over the first few administrations but slowed at later assessments. Given the importance of rapid learning in the EIBI treatment model, it is important to determine if the same steep growth curves are seen with the VB-MAPP.

In summary, there is a small literature evaluating changes in curriculum assessments (e.g., VB-MAPP, ABLLS-R) associated with behavioral treatment. Each study has demonstrated progress in skills targeted during treatment (e.g., Lotfizadeh et al. 2020; Vietze and Lax 2020). However, no studies have examined growth curves on the VB-MAPP as has been done for the ABLLS-R (Myers et al. 2022). In addition, Lotfizadeh et al. (2025) compared scores on the VB-MAPP and the VABS-II at a single point in time during treatment, finding moderate to strong correlations between VB-MAPP Milestones scores and several domains and the ABC of the VABS-II. However, no analyses have been conducted with the current third edition of the Vineland (i.e., VABS-III), and no studies have examined whether correlations between curriculum assessments and adaptive behavior scores change throughout the course of treatment. Thus, additional studies are needed to examine VB-MAPP change scores, growth curves, and the relation between VB-MAPP scores and standardized measures that have been used in prior outcome research.

This study reports changes in scores on the VB-MAPP for a large, diverse sample of children with ASD in EIBI in a routine clinical care setting. We examined correlations between the curriculum measure and domain and ABC scores on the VABS-III at intake and approximately every 6 months throughout treatment. In addition, we examined the variables associated with initial curriculum assessment scores and overall growth scores, as well as growth trajectories throughout treatment similar to the analysis of Myers et al. (2022). Results were examined for all participants and separately by whether the duration of treatment matched the clinical recommendation or was shorter than intended to approximate an intent to treat analysis.

2 | Methods

All assessment, treatment, and discharge activities occurred as part of ongoing care without influence by the primary researcher. Consent for analysis of archival data occurred during the consent-for-services process. This archival study was reviewed and approved by the WCG Institutional Review Board and was conducted in compliance with the ethical guidelines of

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the Behavior Analyst Certification Board and in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

2.1 | Participants

The participants included 282 children with ASD who (a) had been enrolled in an EIBI program for at least 6 months with a discharge date between May of 2023 and December of 2024, (b) were 7.0 years or younger at the start of services, and (c) had at least two administrations of the VB-MAPP, including an assessment dated within approximately 3 months of the first date of treatment (i.e., baseline measurement). There was no exclusion of participants who had previously received ABA services with another provider, nor was there any exclusion based on test scores or any other measure of ability. The third author was the only author with a clinical caseload and oversight of other Board Certified Behavior Analysts (BCBAs) who might have administered or overseen the administration of the VB-MAPP during this time period as part of their clinical duties.

Only 203 (71.9%) of these participants had an available score dated within 7 months of discharge, as clinicians sometimes switched to a different curriculum assessment if the child neared or reached the maximum possible VB-MAPP score during the course of treatment. Analyses of change scores from intake to discharge were restricted to these patients only. See Table 1 for a detailed comparison of participants in the full sample and the subsample with an assessment at discharge. The characteristics were similar with respect to race/ethnicity (diverse), gender distribution (predominantly male), primary language (predominantly English), age at start of services (~3.8 years), fulfillment percentage (~81%), and treatment hours per week (~30). The subsample had a slightly shorter duration of services and a younger age at discharge. Approximately 86% of all participants entered services below the age of 5 years.

All participants had an independently conferred diagnosis of ASD from a qualified provider using either the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; American Psychiatric Association 2013), *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., text rev.; American Psychiatric Association 2022), or the *International Classification of Diseases* (10th ed.; World Health Organization 2016). Results of the diagnostic tools used by the evaluating clinician were available in the medical record for some participants and are provided below to describe the sample. Those results are presented as severity or likelihood descriptors so that similar information can be presented across tools, but this information was not used in the statistical analysis.

TABLE 1 | Participant characteristics for the full sample and the subsample with an assessment conducted at discharge.

Characteristic	Full sample (n = 282)[a]	Assessment at discharge (n = 203)[a]
Ethnicity		
White	76 (27%)	59 (29%)
American Indian or Alaska Native	5 (1.8%)	4 (2.0%)
Asian	48 (17%)	36 (18%)
Black or African American	45 (16%)	28 (14%)
Hispanic or Latino	66 (23%)	48 (24%)
Missing	21 (7.4%)	13 (6.4%)
Multiple ethnicity	13 (4.6%)	10 (4.9%)
Other	8 (2.8%)	5 (2.5%)
Gender		
Female	69 (24%)	45 (22%)
Male	213 (76%)	158 (78%)
Primary language at home		
English	264 (94%)	191 (94%)
Other	14 (5.0%)	9 (4.4%)
Spanish	4 (1.4%)	3 (1.5%)
Age at start (years)	3.83 (1.07)	3.79 (1.04)
Age at discharge (years)	5.57 (1.24)	5.31 (1.09)
Discharge reason		
Premature exit	183 (65%)	139 (68%)
Recommended exit	99 (35%)	64 (32%)
Duration of services (days)	630 (294)	549 (230)
Fulfillment (%)	0.83 (0.12)	0.81 (0.12)
Avg weekly hours	30.9 (5.1)	30.6 (5.3)

[a]n (%); mean (standard deviation).

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A diagnostic severity or likelihood measure was available for 218 of the 282 (77.3%) participants. The Childhood Autism Rating Scale-Second Edition (CARS-2; Schopler et al. 2010) results were available for 115 participants, with five scores in the minimal-to-no symptoms range, 64 scores in the mild to moderate range, and 46 scores indicating severe symptoms. The Autism Spectrum Rating Scales (ASRS; Goldstein and Naglieri 2009) uses the descriptors average (i.e., level of symptoms is not elevated compared to the general population), slightly elevated, elevated, and very elevated (i.e., indicative of an ASD). The ASRS descriptor was available for seven participants, with one described as elevated and six as very elevated. The Autism Diagnostic Observation Scale-Second Edition (ADOS-2; Lord et al. 2012) was available with either standard or modified administration (due to Covid-19) for 96 participants. Of the 29 participants who had an ADOS-2 Module 1–3 comparison score with standard administration, one score was in the low range, 15 were in the moderate range, and 13 were in the high range. Of the 21 participants who had an ADOS-2 Toddler Module with standard administration, four scores were in the mild-to-moderate range of concern and 17 were in the moderate-to-severe range of concern. All 46 participants who had an ADOS-2 with a modified administration were scored as definite differences.

2.2 | Model of Treatment

The organization provides EIBI treatment services in six states in the United States. Approximately 95% of children are served in centers and 5% are served in their homes. Services are paid for by state Medicaid agencies or private health insurance. Based on the results of a functional behavioral assessment, curriculum assessment, adaptive behavior assessment, and family collaboration interviews, the BCBA requests authorization (i.e., request to the funder to allow and pay for services at a certain level for a specified duration of time) for direct treatment services for the child, supervision and protocol design, and family guidance sessions for the parents/caregivers. The typical length of the authorization period was 6 months, but this interval varied by funding source and could have been shorter than 6 months or as long as 1 year.

The request for services varied based on individual need but generally remained consistent with the Council of Autism Service Providers (CASP) guidelines (Council of Autism Service Providers 2024), with approximately 30–40 h per week of direct treatment for the child, as well as 15%–20% supervision by the BCBA. In addition, the request for services included at least two contacts for family guidance per month. The process was repeated if services continued beyond the initial authorization period, as the generally accepted standard of care for the duration of treatment is two or more years (Council of Autism Service Providers 2025). The number of treatment hours per week may have been lower in the first authorization for a very young child or in the later authorization period(s) for a child of any age as part of a titration in services in preparation for transition to school. See Table 2 for the authorized weekly direct treatment hours per authorization interval (i.e., the approved duration of services by the funder). The majority of children received at least 30 hours per week of direct treatment in every authorization period, and authorization fulfillment (i.e., the percentage of recommended direct treatment hours that actually occurred) ranged from 81.9% to 93.3% across authorization periods.

Children were typically scheduled to attend behavioral treatment sessions for 6–8 h per day, 5 days per week. Treatment was provided in a one-to-one ratio in both individual and small group instruction using structured and natural environment behavioral teaching procedures designed to be fun, engaging, and play-based (Geiger et al. 2012) and consistent with an assent-based model of care (Breaux and Smith 2023). Parents typically attended parent guidance sessions twice per month to observe sessions, collaborate on goal and intervention selection, learn teaching strategies, and receive consultation for newly emerging concerns (e.g., sleep issues, new topographies of challenging behavior, safety concerns).

2.3 | Measures

2.3.1 | Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP).

The VB-MAPP Milestones measures 16 domains targeting 170 key developmental social and language milestones in three developmental levels. Each item can be scored as 0, ½, or 1 point, resulting in a total possible score of 170. See the prior section for additional information about the psychometric properties of the VB-MAPP. The Board Certified Behavior Analyst (BCBA) BCBA assigned to the patient was responsible for completing the VB-MAPP, developing the treatment plan and programs, and supervising all aspects of service delivery and family services. Although every effort is made to ensure stability in patient assignments, the BCBA may have changed at some point in the course of treatment for some patients. Many BCBAs are taught to administer and score this assessment as part of their master's-level fieldwork experience prior to becoming certified, but any BCBAs who were not previously trained on this assessment were trained by their clinical director on administration, scoring, and interpretation of the VB-MAPP in their center as part of onboarding and training for their position. In addition, the clinical director provided close oversight of all assessments, regardless of prior training, until a performance criterion was met for the quality of assessment and treatment reports. The BCBA completed the VB-MAPP over one or more observations. Although the BCBA may have completed additional components of the VB-MAPP (e.g., Barriers Assessment), only the Milestones total score was included in this analysis. A request for treatment authorization was made for every authorization interval until discharge, and the results of the VB-MAPP were used to guide programming for that treatment authorization interval.

2.3.2 | Vineland Adaptive Behavior Scales-Third Edition (VABS-III), Caregiver Rating Form, Comprehensive

The VABS-III (Sparrow et al. 2016), Caregiver Rating Form is a norm-referenced parent questionnaire assessing adaptive skills that can be used to evaluate change over time. The comprehensive form includes 381 items that the caregiver scores for how often their child performs skills (0 = never, 1 = sometimes, 2 = usually or often). In addition to the summary measure (i.e.,

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Treatment intensity and authorization fulfillment per authorization period.

	AP1 ($N = 282$) (%)	AP2 ($N = 265$) (%)	AP3 ($N = 179$) (%)	AP4 ($N = 113$) (%)	AP5 ($N = 51$) (%)	AP6 ($N = 27$) (%)	AP7 ($N = 12$) (%)	AP8 ($N = 6$) (%)	AP9 ($N = 2$) (%)	AP10 ($N = 1$) (%)
30–40	70.9	66.8	68.2	64.6	66.7	59.3	83.3	83.3	100.0	100.0
20–29	24.8	26.8	26.8	24.8	19.6	25.9	0.0	0.0	0.0	0.0
Below 20	4.3	6.4	5.0	10.6	13.7	14.8	16.7	16.7	0.0	0.0
Avg. fulfillment %	**81.9**	**82.2**	**84.8**	**84.9**	**86.1**	**87.4**	**93.2**	**93.3**	**90.0**	**93.0**

Note: The first three rows indicate the percentage of total patients in that time period authorized to receive treatment hours in the specified range. Due to rounding, these percentages may not add up to exactly 100%. The fourth row (bolded values) indicates the average percentage of authorized hours that were provided across all children for that authorization period.
Abbreviation: AP = Authorization Period.

ABC Standard Score), Communication, Socialization, Daily Living Skills, and Motor Skills domain standard scores are calculated. Reliability and validity indices are high (i.e., mean internal consistency for the ABC = 0.99 across age groups; test-retest and interrater reliability coefficients in the excellent range; domain raw scores align with expected developmental trends in adaptive behavior; Sparrow et al. 2016). The BCBA assigned to the patient was responsible for sending the parent a link to the electronic version of the VABS-III Caregiver Rating Form. The parent completed the VABS-III every 6–12 months as required by the funder. All instructions and item presentation were automated by the test publisher, and all responses were completed in, and scored by, the test publisher's Q-Global platform. A detailed summary report was provided to the BCBA, who incorporated the results into the assessment report and goal selection process.

2.4 | Discharge Recommendations and Categories

Near the end of each authorization period, the BCBA made a recommendation for (a) continuation of EIBI services or (b) discharge from services. Continuation of services was recommended when scores were more than 1 standard deviation below the mean on a school readiness assessment or the VABS-III, or challenging behavior continued to occur at moderate or severe levels during the prior month. Discharge to school services was recommended when a child scored at a standard score of 85 or higher on a school readiness assessment and the VABS-III and had no or minimally occurring challenging behavior during the prior month. In addition, if one of the three criteria were not met for a school-aged child, but the team felt that the maximal benefit of EIBI had been achieved, discharge to school services was recommended. Families then made the ultimate decision whether to continue with services as recommended, resulting in a discharge categorization of (a) recommended exit or (b) premature exit (i.e., prior to a clinical recommendation to discontinue treatment). This variable was examined because prior analyses (LeBlanc et al. 2025) have identified differential outcomes by discharge category.

2.5 | Statistical Analyses

Bayesian correlations were computed between the VB-MAPP total score and VABS-III domain and total (i.e., ABC) standard scores at every assessment interval at which both measures were administered within a 6-week window. In addition, a Bayesian correlation was calculated between the VB-MAPP total score and participant age at intake. All correlations were estimated with the *BayesFactor* R package (version 0.9.12-4.7; R. Morey and Rouder 2026), and each is summarized by the posterior median and equal-tailed 95% credible interval.

We estimated a Bayesian multilevel growth model to examine how VB-MAPP scores change over time since the start of services. Each child contributed repeated assessment scores (typically spaced approximately 6 months apart), and observations were nested within the child. The outcome variable was the total VB-MAPP score at each assessment. Time since start of service (in years) was modeled with both a linear and quadratic term to allow for potential nonlinear growth. The model also included mean-centered child age at start of service, child gender, ethnicity, discharge reason, and interactions between time and both age and discharge reason. Random intercepts and random slopes for time were included at the child level, allowing children to differ both in their initial VB-MAPP scores and in their individual rates of growth.

We placed broad, regularizing priors on the VB-MAPP score scale rather than default flat priors. The intercept (i.e., predicted score at the start of services) received a Normal (35, 30) prior centered near the observed mean intake score; the linear and quadratic time coefficients received Normal (0, 30) and Normal (0, 20) priors, respectively; the remaining regression coefficients received Normal (0, 15) priors; the child-level random-effect standard deviations received half-Normal (0, 30) priors; the random-effect correlation received a uniform LKJ (1) prior; and the residual standard deviation received a half-Normal (0, 15) prior. A prior-predictive check confirmed that these priors placed the bulk of prior-predicted scores within the VB-MAPP's 0–170 range without concentrating them in an implausible range. Models were estimated in Stan (via *cmdstanr*) with four chains of 4000 iterations each (1000 warmup) and a target acceptance rate of 0.95. Convergence of the VB-MAPP growth models was satisfactory (all R-hat ≤ 1.004, bulk effective sample sizes above 500 for every parameter and above 1300 for every reported coefficient, and no divergent transitions). Posterior predictive checks supported the Gaussian likelihood as a reasonable approximation, although a small fraction of predictive mass fell outside the VB-MAPP's range (1.4% above 170 and 2.9% below 0), which is noted as a limitation of the unbounded

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

likelihood. No frequentist null-hypothesis significance tests or *p* values were reported. Each growth-model effect is summarized by its posterior mean and 95% credible interval.

This Bayesian approach provides more intuitive and informative results compared to traditional methods. Instead of relying on a single estimate and a confidence interval that can be difficult to interpret (R. D. Morey et al. 2016), Bayesian inference gives a range of plausible values (known as credible intervals) that capture the uncertainty around findings. All growth-model effects are summarized using posterior means and 95% credible intervals (CrI; comparable to 95% *confidence* intervals derived from frequentist approaches), which represent the range within which the true parameter lies with 95% posterior probability. A relatively narrow credible interval that does not overlap zero provides stronger confidence in the findings. This modeling strategy (i.e., a quadratic term) mirrored the growth-curve approach in Myers et al. (2022).

Categorical variables were coded using sum-to-zero (effects) contrasts, such that each coefficient reflects the difference between a category and the overall average of the factor (i.e., the grand mean), rather than a difference from a single reference group. When there are only two categories, the coefficient for one category represents its deviation from the mean of both groups, and the other category's effect is equal in magnitude but opposite in sign. In this case, the total difference between the two groups is twice the reported coefficient (J. Cohen et al. 2003).

We tested an alternative model that included average weekly utilized hours for each assessment period as a proxy to "dose", but a comparison of the expected log predictive density (ELPD, a measure of how well the model is expected to predict new data) using leave-one-out cross-validation (LOO-CV; a method for estimating out-of-sample predictive accuracy) favored the simpler model without dose (ELPD difference = −4.84; SE = 2.86), and the dose coefficient was small and uncertain (0.008, 95% CrI: [−0.25, 0.27]). This may reflect the limited variation in weekly hours across assessment periods. Because utilization data were available for only a subset of assessments, this comparison used 689 of the 994 observations, and 70 children contributed only a single assessment to it. Cross-validation is least reliable when many children contribute few observations, so the ELPD comparison was treated as suggestive rather than decisive, and rest the decision primarily on the near-zero dose coefficient. Given this, we opted to report results from the simpler model without weekly hours as a predictor.

All statistics were calculated using the R Statistical language (version 4.5.1; R Core Team 2024) on macOS 26.1 with the *brms* package (version 2.21.0; Bürkner 2017). Other primary packages used included *easystats* (version 0.7.3; Lüdecke et al. 2022), *modelsummary* (version 2.1.1; Arel-Bundock 2022), *bayestestR* (version 0.14.0; Makowski et al. 2019), *marginaleffects* (version 0.21.0; Arel-Bundock et al. 2024), and *ggplot2* (version 3.5.1; Wickham 2016).

3 | Results

Figure 1 depicts a heat map of the Bayesian correlations between the VB-MAPP total score and VABS-III ABC (i.e., Total Score) and domain standard scores for assessments conducted within 6 weeks of each other throughout the course of treatment. The sample size for each correlation is specified within each cell. Each participant contributed between two and seven VB-MAPP administrations (median = 3), with consecutive administrations spaced a median of 178 days apart; A1 denotes each participant's first VB-MAPP administration, A2 the second, and so on. Additional descriptive statistics about the number and timing of VB-MAPP administrations are provided in Supporting Information S1: Table S1. This analysis attempted to determine whether the findings of Lotfizadeh et al. (2025)—that is, moderate to strong correlations between VB-MAPP and VABS-II ABC and certain domains at a single point in time—were also evident for the VABS-III and whether the relation between the two measures changed throughout the course of treatment. Positive correlations are depicted in green, with stronger correlations depicted as darker colors. The sample size decreased with each subsequent correlation, with only four participants included in the final correlations (i.e., A5). Given the small sample size ($n = 4$), A5 is not interpreted.

The correlations were consistently strongest between VB-MAPP scores and the VABS-III Communication domain standard score (median range, $r = 0.50$ to 0.68), followed by the ABC standard score (median range, $r = 0.40$ to 0.63), providing a replication of Lotfizadeh's et al.'s evidence of validity for the VB-MAPP as a communication-focused curriculum assessment. The correlations between the VB-MAPP and the Motor Skills domain were always the lowest at any given time, ranging from negligible at the first assessment (i.e., the 95% credible interval [CrI] overlapped zero) to low to moderate correlations at the second, third and fourth assessments (median range, $r = 0.09$ to 0.45). These low correlations are consistent with instrument content, given that the VB-MAPP Milestones assessment contains no motor items—a divergent pattern also reported by Lotfizadeh et al. (2025). For the ABC and most domains, the correlations were stronger at each subsequent assessment interval (i.e., A2 correlation > A1, A3 correlation > A2, A4 correlation > A3). That is, throughout treatment as both VB-MAPP and VABS-III scores increased and as more participants exited services, these measures became more highly correlated. Standard scores rose over the treatment period for the ABC and for the Communication, Socialization, and Daily Living domains (3.8–5.3 points per year), whereas the Motor Skills domain changed little; the model and estimates are reported in the Supporting Information S1 (Figure S1).

Two features of these data invite caution. First, because both measures were administered at broadly similar points in treatment, the correlations could, in principle, reflect shared elapsed time rather than a substantive association. However, repeating each correlation with an adjustment for time since the start of services left the estimates essentially unchanged (a shift of ≤ 0.01 at every domain and wave), which indicates that the associations are not a result of linear dependence on elapsed time. Second, the apparent strengthening of the correlations across successive assessments coincides with a change in sample composition, given that children with later assessments are, by definition, those who remained in services longer. Restricting the analysis to a balanced cohort of 81 children with a valid Total-Score pair at every wave from A1 to A3,[1] the point

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



FIGURE 1 | Bayesian median correlation coefficients between Vineland scores and VB-MAPP total score with 95% credible intervals and sample sizes.

estimates increased from A1 to A3 in every domain (e.g., the ABC rose from $r = 0.26$ at A1 to $r = 0.53$ at A3; estimates within this restricted cohort are lower than the full-sample values in Figure 1, as expected from the smaller and more homogeneous sample), which makes changing membership an unlikely complete explanation. Both analyses are treated as sensitivity checks; the general pattern survives these adjustments, but neither identifies within-child strengthening of the association.

The remaining tables and figures explore change on the VB-MAPP throughout treatment and the variables associated with that change. Figure 2 (left panel) depicts the distribution of individual participant VB-MAPP scores at intake to treatment services for all participants. This analysis allows examination of the heterogeneity of the sample with a focus on individual participant data as well as measures of central tendency, as measures of central tendency can sometimes obscure the complexity of a data set. The distribution is heavily positively skewed (i.e., the majority of children had very low scores with a few higher scores, making the mean higher than the median, which was in turn higher than the mode) at intake, demonstrating the significant needs evident for this sample of children. Age was positively correlated with VB-MAPP score at intake ($r = 0.356$, 95% CrI [0.255, 0.452]; $n = 282$). The middle and right panels depict the distribution of scores at intake and discharge for the subset of participants ($n = 203$) with scores available at both time points. The distribution of scores at intake does not differ from the distribution at intake for the entire sample (i.e., also heavily skewed, similar measures of central tendency), suggesting that the sample for whom we have both intake and discharge data did not differ significantly from the entire sample. However, the distribution at discharge (i.e., after treatment) shifted significantly, with the mean (103), the median (112), and the area where 50% of the scores fall (denoted by the shaded box) having no overlap with the intake distribution. After treatment, the distribution is not simply shifted upward; rather, the shape of the distribution is altered by the end of treatment.

Figure 3 depicts total change scores (i.e., difference between intake and discharge measurements) on the VB-MAPP for each participant, arranged from lowest to highest change score and by discharge category (i.e., recommended exit, premature exit). This analysis allows examination of individual participant data as well as measures of central tendency to depict the heterogeneity evident in treatment progress and whether change scores differed by discharge category. The full range of possible change scores was represented in the sample, and the mean total change score was 69.6 (median 69.5). The total change scores were, on average, higher for children who remained in service until recommended discharge ($M = 84.9$ points) than for those who exited services prematurely ($M = 62.6$ points);

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analysis Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

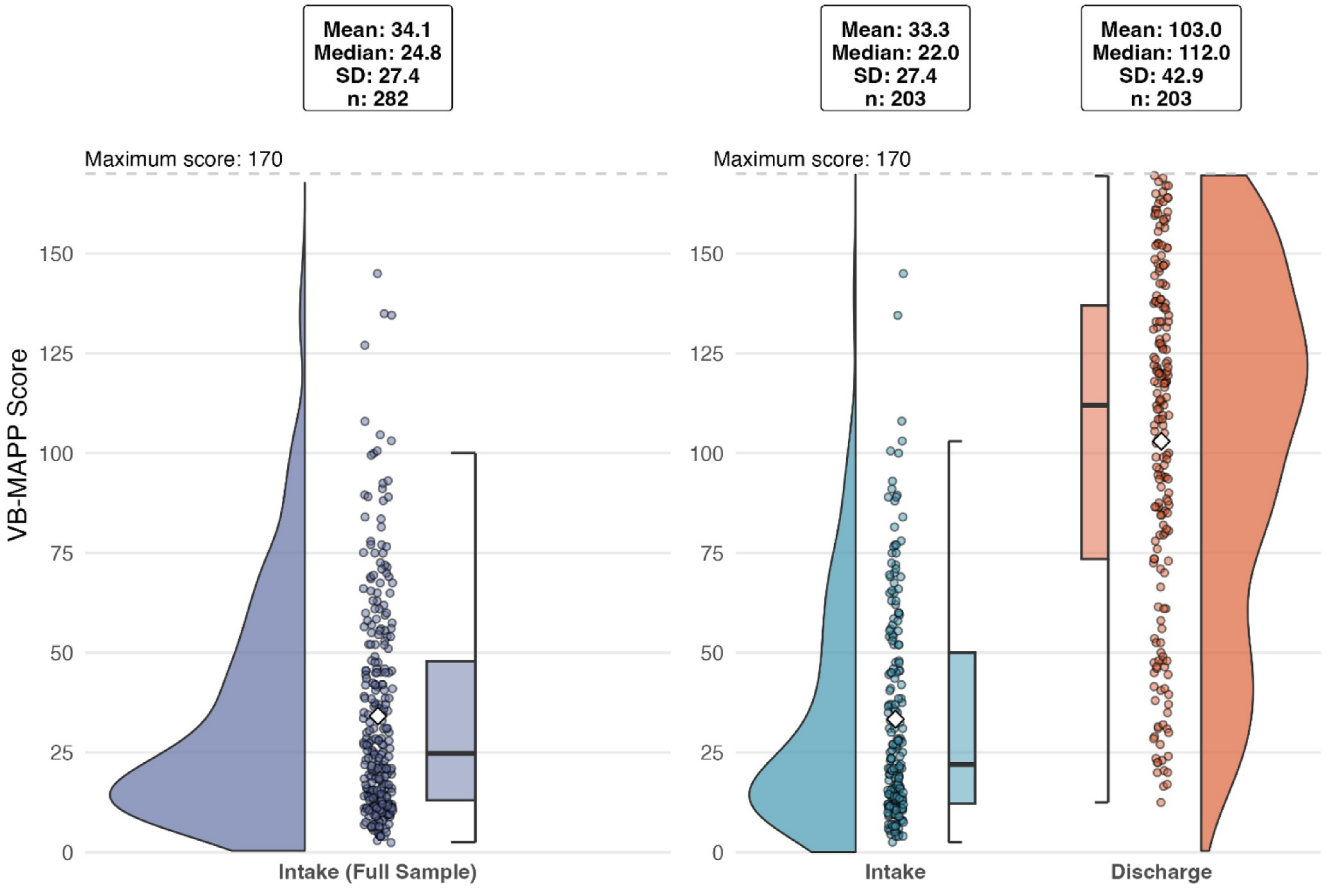


FIGURE 2 | Distribution of VB-MAPP scores at intake for all participants (n = 282) and at intake and discharge for the subset with both measurements (n = 203). The boxes indicate where 50% of the data points lie. The horizontal lines inside the boxes indicate the median score. The open diamonds indicate the mean score.

however, some children who exited prematurely experienced more change than some children who had a recommended discharge.

Figure 4 depicts a growth-curve analysis representing data for each individual participant. This analysis allows examination of whether different rates of growth are seen at different times throughout the treatment process, as has been noted for the ABLLS-R. The VB-MAPP scores increased over time, with evidence for a nonlinear pattern of growth: gains were strongest early in the service period and then decelerated. The average predicted score at intake was 36.1 (95% CrI [33.1, 39.1]), consistent with the observed mean intake score of 34.1 (the model intercept, the predicted score at the start of services for a child at the covariate grand mean, was 43.19; Table 3). On average, scores increased by 74.81 points per year (95% CrI [70.68, 78.90]) early in the service period, but this growth slowed over time, as indicated by a negative quadratic term (−10.40, 95% CrI [−12.21, −8.64]). There was substantial variability in both baseline scores and growth trajectories across individuals. The standard deviation of random intercepts was 23.99 (95% CrI [21.79, 26.38]), and the standard deviation of time slopes was 22.42 (95% CrI [19.36, 25.71]); that is, children differed widely both in the skills they brought to treatment and in how quickly they gained skills, so the average trajectory should not be read as describing any individual child. The correlation between intercepts and slopes was 0.43 (95% CrI [0.28, 0.57]), indicating that children with higher initial scores tended to improve faster over time. The residual standard deviation, reflecting within-person variation not explained by the model, was 10.97 (95% CrI [10.22, 11.75]). Central tendency statistics for change scores can be found in Supporting Information S1: Table S2 and reveal the same pattern, with the largest change occurring from A1 to A2 and from A2 to A3. Mean change scores were positive at every interval, indicating that participants continued to experience skill development and growth throughout treatment.

Table 3 provides the Bayesian regression results. This analysis was conducted to examine whether children's starting scores differed by age at the start of services, gender, race/ethnicity, and discharge category, and whether rates of growth differed by age and by discharge category, while accounting for the overall growth pattern described above. Older children (at start of services) started with higher scores, and each year increase in age at start of services was associated with 10.75 higher points at intake (95% CrI [7.90, 13.54]). The interaction between age and time was small and uncertain (−1.15, 95% CrI [−4.03, 1.73]), so no difference in the rate of improvement by age is interpreted. Males scored 3.99 points below the overall gender mean (95% CrI [−7.45, −0.53]); under sum-to-zero coding this corresponds to a male–female difference of roughly eight points. White children scored 10.50 points above the grand mean (95% CrI [4.21, 16.88]), whereas most other ethnicity contrasts were small and uncertain. Children who exited prior to a clinical recommendation scored

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

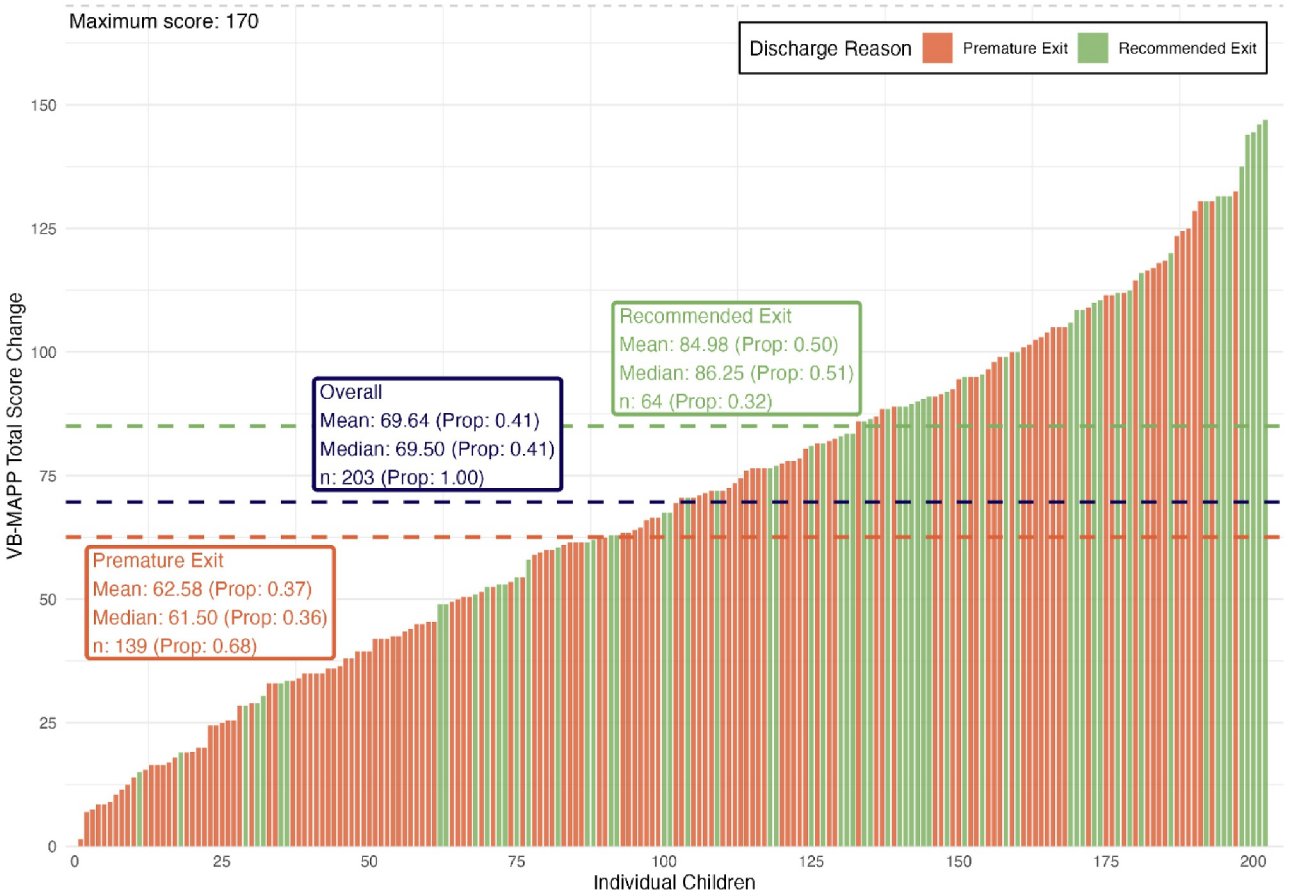


FIGURE 3 | Individual participant improvement on the VB-MAPP as total change scores.

8.04 points below the discharge mean (95% CrI [−11.21, −4.86]) and had a slower rate of improvement over time (−7.28, 95% CrI [−10.68, −3.95] for the time × discharge interaction).

4 | Discussion

Recent studies have found moderate to strong correlations between VB-MAPP scores and prior editions of the VABS (Lotfizadeh et al. 2025) and positive change scores on the VB-MAPP for children in behavioral treatment (Vietze and Lax 2020; Lotfizadeh et al. 2020). However, the literature is sparse and has only examined change scores and correlations from intake to a single point in time (i.e., discharge or 2 years of therapy) rather than throughout the course of therapy. The current study extends the literature on curriculum assessment scores in routine clinical care for autistic children in several ways. First, we explored correlations between scores on the VB-MAPP and the VABS-III throughout the course of treatment. Second, we examined changes in VB-MAPP scores and growth trajectories throughout the course of treatment. Third, we examined the relationship between age at intake and initial scores, and whether change scores differed by discharge category and several other variables in a regression analysis.

Our findings add to the validity evidence for the VB-MAPP. Lotfizadeh et al. (2025) found moderate to strong correlations between the VB-MAPP Milestones score and the Communication, Socialization, and Daily Living domains and the ABC standard score, but a lower correlation with the Motor Skills domain. These findings were interpreted as evidence of convergent and divergent validity for the VB-MAPP. Their analyses used an older version of the Vineland Interview form (i.e., VABS-II) at a single assessment interval during treatment services. Our own analyses focused on the current edition (i.e., VABS-III) Caregiver Rating Form and included standard scores rather than raw scores. However, the results were similar in that the strongest correlations were evident for the Communication domain, followed by the ABC, then the Socialization and Daily Living domains, with relatively low correlations with the Motor Skills domain. Thus, our findings replicate those of Lotfizadeh et al. (2025), providing additional validity evidence for the VB-MAPP using the most recent version of the Vineland. In addition, correlations between the two measures generally increased across assessment intervals during the first 2 years of treatment as each of the scores increased. These findings indicate that the relationship between progress in the behavioral curriculum during treatment and standardized tests of adaptive behavior was strong and that this strong relationship has now been demonstrated with multiple versions and editions of the Vineland and using both raw scores and standard scores. These increasing correlations throughout treatment also suggest that future studies examining correlations between these two measures should specify the timing with respect to treatment (e.g., no treatment, during treatment, post treatment).

Previous studies examined change scores on the VB-MAPP (e.g., Lotfizadeh et al. 2020; Vietze and Lax 2020) from intake to discharge but not throughout treatment, limiting our

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

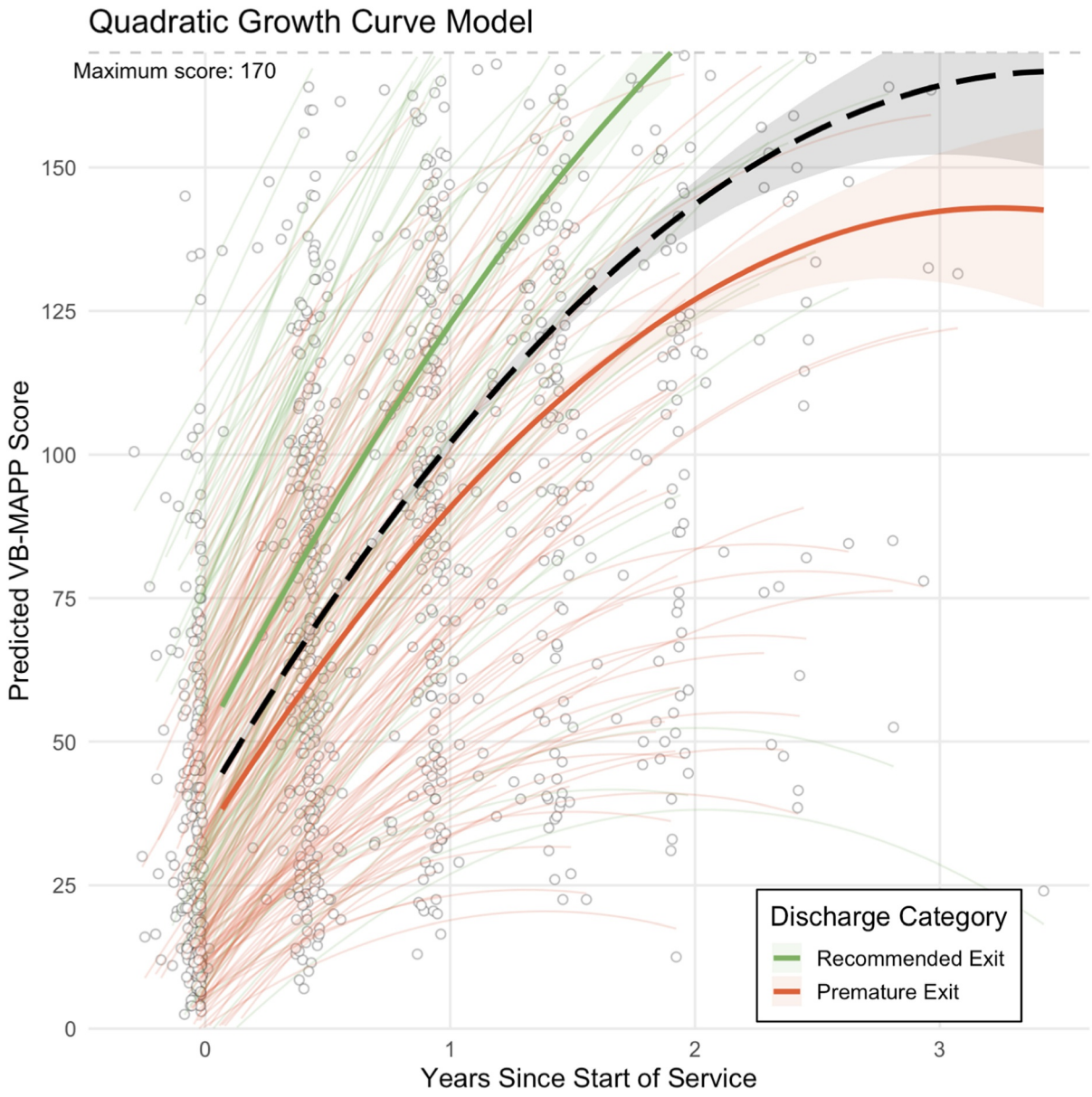


FIGURE 4 | Quadratic growth curve analysis (n = 282). Nearly all intake assessments (280 of 282) were conducted prior to the beginning of treatment, which is denoted as 0 on the x-axis. For most patients the start of treatment occurred weeks to a few months after the assessment was completed, thus their first data points fall to the left of the 0 mark; for two patients the first assessment occurred shortly after the start of treatment, placing their first data point just to the right of 0. The Recommended Exit curve is drawn only where its model-implied mean lies within the VB-MAPP's range (see Statistical Analysis for more information).

understanding of growth patterns over time. Lotfizadeh et al. reported average change scores of 69 points (converted from their percent score calculation) over a span of 2 years of treatment, however, they excluded all patients with high intake scores. In comparison, the mean change score for all participants with intake and discharge scores in the current study was also approximately 70 points, but that change occurred over a shorter span of time (i.e., about 1.5 years) and did not exclude higher scoring patients at intake, who experienced slower progress. Similar to Lotfizadeh et al. (2020), an analysis of the individual change scores (see Figure 3) revealed a full continuum of change scores, with greater average change for those with a recommended discharge (M = 85) compared to those who exited prematurely (M = 62.6). The individual participant analyses in both studies demonstrate the heterogeneity of skill presentation at the onset of treatment as well as in response to treatment across time.

Like (LeBlanc et al. 2025), this analysis also revealed differential results for those who remained in service for the recommended duration versus those elected to pursue school rather than continuing in services. A relatively high percentage of families chose to pursue school, on average at about 1.5 years of treatment services. The decision to pursue public school rather than private health care services could occur for many reasons, including financial considerations, transportation convenience, societal expectations for age at entry to school, preference for having their child with same age peers during the day, and

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Bayesian regression summary.

VB-MAPP (quadratic)							
Family: Gaussian | Observations: 994							
			95% CrI		R-	ESS	ESS
Term	**Estimate**	**SE**	**Lower**	**Upper**	**hat**	**(bulk)**	**(tail)**
Residual							
sigma (residual)	10.97	0.39	10.22	11.75	1.00	2947	5877
Random effects (child level)							
SD (intercept)	23.99	1.18	21.79	26.38	1.00	4768	7936
SD (time since start of service)	22.42	1.61	19.36	25.71	1.00	3113	6503
Cor (intercept, time since start of service)	0.43	0.07	0.28	0.57	1.00	4733	6530
Fixed effects (μ)							
(Intercept)	43.19	2.36	38.54	47.83	1.00	5138	7126
Time since start of service (years)	74.81	2.09	70.68	78.90	1.00	9491	9265
Time since start of service (years)^2	−10.40	0.92	−12.21	−8.64	1.00	13,367	9577
Age (centered)	10.75	1.43	7.90	13.54	1.00	4094	6424
Gender—Male	−3.99	1.76	−7.45	−0.53	1.00	4529	6515
Ethnicity—White	10.50	3.19	4.21	16.88	1.00	4376	6588
Ethnicity—American Indian or Alaska Native	5.70	8.25	−10.45	21.87	1.00	7657	8166
Ethnicity—Asian	−3.14	3.73	−10.42	4.17	1.00	4875	6866
Ethnicity—Black or African American	−6.09	3.72	−13.37	1.21	1.00	5024	6412
Ethnicity—Hispanic or Latino	−2.01	3.30	−8.54	4.50	1.00	4451	6262
Ethnicity—Missing	1.21	4.93	−8.45	10.80	1.00	5189	6929
Ethnicity—Multiple ethnicity	−6.57	5.72	−17.64	4.76	1.00	6458	7677
Discharge—Premature exit	−8.04	1.62	−11.21	−4.86	1.00	3355	6750
Time since start of service (years) × age (centered)	−1.15	1.47	−4.03	1.73	1.00	6443	8056
Time since start of service (years) × discharge—Premature exit	−7.28	1.69	−10.68	−3.95	1.00	7546	8458

Note: Factors are effects-coded (sum-to-zero), so each contrast is a deviation from the grand mean; the omitted contrasts are Gender—Female, Ethnicity—Other, and Discharge—Recommended exit.
Abbreviations: Cor = correlation between the random intercepts and slopes, SD = standard deviation of the child-level random effects.

belief that their child has experienced the maximum benefit of treatment, among others. Lower VB-MAPP scores and comparatively slower growth curves (on average) were evident for children who exited prior to recommendation compared to those who were recommended to discharge. No conclusions can be drawn about causality, but future studies might examine the variables that influence the decision to exit or remain in services until recommended discharge.

Vietze and Lax (2020) reported a mean change score of 44 points on the VB-MAPP for toddlers in a low intensity behavioral early intervention program. This mean change score is lower than the mean for the current sample and Lotfizadeh et al. (2020) (i.e., $M = 69$ points); however, the duration of the program was less than 1 year, and the average age at discharge was younger than the average age at intake for the current sample. Our own regression analysis indicated that children starting services at a younger age had lower scores at intake and that rates of improvement during treatment did not differ credibly across the age range, meaning younger children at the start of services were gaining skills during a developmental window in which, absent effective intervention, the adaptive gap has been reported to widen (Casula et al. 2024; Szatmari et al. 2015), supporting the critical role of early identification and intervention. Taken together, these findings and those of MacDonald et al. (2014), suggest that when children are identified early and receive treatment services, the effects of treatment may be larger and may be achieved at shorter treatment durations. However, behavior analysts are cautioned to remember that the VB-MAPP is not a standardized measure, and change scores cannot be interpreted against a success threshold the way scores on standardized measures can (i.e., standard scores, minimum clinically important difference scores). That said, data from multiple studies from routine care settings could eventually be used to establish benchmarks and ranges for expected change scores at different points in treatment.

The growth-curve analysis parallels that of Myers et al. (2022) and similarly reveals steeper growth earlier in treatment. The largest change scores occurred early in treatment (i.e., between A1 and A2 and between A2 and A3), with a lower magnitude of change later in treatment. These findings are consistent with the

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

purpose of EIBI to accelerate learning rates in order to close developmental gaps (MacDonald et al. 2014). However, on average, improvement occurred at every assessment interval, even for those who were in treatment for the longest durations and who had slower growth rates throughout treatment. Thus, it would be inappropriate to deny access to treatment services for those with slower response rates and/or longer treatment durations who are still making progress on important curriculum targets. It is important to note that a ceiling effect was evident as some children achieved near-maximum scores and then transitioned to a different assessment during the course of ongoing treatment. This suggests that future editions of the VB-MAPP would benefit from expansion of curriculum targets at higher developmental levels. In addition, this is only the first study of growth curves on the VB-MAPP, and additional studies should be conducted to determine if these findings are replicated.

Finally, in addition to age at the start of services and discharge category (discussed above), two other variables were predictive of VB-MAPP scores. Males, who were the majority of the sample (i.e., over 75%), scored below the overall gender mean at any given point in treatment (a male-female difference of roughly eight points), suggesting that females in the sample (i.e., below 25%) had comparatively higher scores. The model did not include interactions between time and either gender or race/ethnicity, so these coefficients describe overall score level and are not informative about differences in the rate of progress during treatment. In addition, participants identified in the White race/ethnicity category (representing less than 30% of the sample) also had higher scores on average, with few differences for any other category. The predictive value of these two demographic variables reinforces the importance of the call for more detailed reporting of demographic variables in behavior analytic research (Jones et al. 2020). However, it is likely that associated unmeasured third variables account for the observed effects rather than the measured demographic variables themselves (e.g., resources and broader system inequities rather than skin pigment; gender-associated differences in co-occurring medical conditions and diagnostic patterns rather than direct X or Y chromosome effects). These findings highlight the importance of equitable access to EIBI and the dangers that would arise from access or dosage decisions being based on these types of treatment-response predictive analyses. Rather, these types of analyses should be used to help understand when a longer duration of treatment might be required to achieve good outcomes.

All of these findings should be interpreted in the context of the limitations of retrospective analyses in routine clinical care. Retrospective studies, such as this one, are only able to examine the data that exist in the record to describe variables and covariates without a direct manipulation or control for those variables. Rather than firm *causal* conclusions about variables, conclusions are limited to variables that are *associated* with differences in outcome. For example, although age at intake covaried with initial scores, these data indicate correlation rather than causation. Prospective studies typically control parameters of test administration (i.e., intervals between assessments, selection of assessments, assessors who are not also the treatment provider, secondary observers) as well as treatment (i.e., treatment intensity, duration of treatment, no prior exposure to treatment). In routine clinical care, these variables are heavily influenced by individual provider choices, generating what has been referred to as real world data with various limitations. Total change scores could not be calculated when the BCBA switched assessments as treatment progressed, usually due to ceiling effects with the VB-MAPP. Children who made the most rapid progress in treatment may have been eliminated from the total change scores analysis, resulting in an underestimate of change scores. In addition, many different BCBAs administered the assessments across participants, and potentially for the same participant over time. Assessor-related variables (e.g., bias, variability in administration) likely account for some portion of the observed variability in scores over time. In addition, treatment intensity recommendations were directly based on the measures examined in the study rather than held constant, so the effects of treatment intensity as an independent variable on progress cannot be examined. The sample also included participants with potential prior exposure to treatment at some unknown intensity and duration. Finally, the duration of treatment was shorter than intended for some participants (i.e., premature exit) and was, on average, shorter than the duration in controlled studies (approximately 18 months compared to 2 years or more). However, these conditions represent routine clinical care as encountered by many children with autism in the United States with health care-funded treatment services.

These limitations are managed, though not eliminated, by several features of the study's design. Members of the investigative team are employed by the organization that provided treatment, and we structured the analysis to limit the influence of that affiliation on the results: all assessment, treatment, and discharge decisions occurred as part of routine care without direct influence by the research team (see Methods); inclusion required only two VB-MAPP administrations rather than a complete battery of every measure, so children with incomplete records were retained (the smaller samples for the VABS-III correlations reflect this choice); children were followed to their actual point of discharge rather than to a fixed study duration; and we report individual-level distributions and change scores alongside model-based averages, so the full range of outcomes, including children who progressed slowly, remains visible.

In summary, the current findings contribute to our understanding of the changes observed on the VB-MAPP in a routine clinical care setting for a large, diverse sample of young children with ASD in center-based EIBI services. Lower scores were associated with male gender and premature exit, and children who were older at intake began with higher scores but did not differ credibly in their rate of improvement; however, children remaining in services for multiple treatment authorizations continued to experience positive gains, on average, at every assessment interval. As more published studies provide detailed data on changes in curriculum assessments, benchmarks for expected change scores may eventually be established to provide a comparison for providers to evaluate the relative effects of their treatment services. In addition, the correlations between the VB-MAPP scores and standardized scores on the VABS-III provide evidence that treatment gains are reflected in measures other than the assessment that directly guides programming.

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Contributions

Linda A. LeBlanc and Brent A. Kaplan conceptualized the study and wrote the manuscript. Brent A. Kaplan conducted all statistical analyses and created all figures and tables. Kelsie Thompson, Alec Sweatman, and Charna Mintz reviewed and edited the manuscript. Kelsie Thompson and Alec Sweatman managed the data set and retrieved missing data.

Acknowledgments

The authors thank Shanky Balani and Hailey Rivera for their assistance with compiling the data set. Text descriptions of the treatment model and diagnostic confirmation information are similar to other manuscripts on treatment outcomes from this organization.

Funding

The authors have nothing to report.

Ethics Statement

This study was reviewed and approved by the WCG Institutional Review Board and has been conducted in compliance with the guidelines of this board as well as ethical guidelines of the Behavior Analyst Certification Board and in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Consent

Consent for use of data was obtained at the outset of services.

Conflicts of Interest

Linda A. LeBlanc and Charna Mintz have equity in and are employed by Action Behavior Centers. Brent A. Kaplan is a statistical consultant to Action Behavior Centers. Kelsie Thompson and Alec Sweatman are employed by Action Behavior Centers. No other potential COIs are noted. The study is an evaluation of the services provided by the organization that employs Linda A. LeBlanc, Kelsie Thompson, Alec Sweatman, and Charna Mintz.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

[1] The Motor Skills domain was available for 79 to 80 of these children rather than all 81.

References

Alvares, G. A., K. Bebbington, D. Cleary, et al. 2020. "The Misnomer of 'High Functioning Autism': Intelligence Is an Imprecise Predictor of Functional Abilities at Diagnosis." *Autism: The International Journal of Research and Practice* 24, no. 1: 221–232. https://doi.org/10.1177/1362361319852831.

American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. https://doi.org/10.1176/appi.books.9780890425596.

American Psychiatric Association. 2022. Diagnostic and Statistical Manual of Mental Disorders. 5th ed., text rev. https://doi.org/10.1176/appi.books.9780890425787.

Arel-Bundock, V. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103, no. 1: 1–23. https://doi.org/10.18637/jss.v103.i01.

Arel-Bundock, V., N. Greifer, and A. Heiss. 2024. "How to Interpret Statistical Models Using Marginal Effects in R and Python." *Journal of Statistical Software* 55, no. 2: 31. https://marginaleffects.com.

Behavior Analyst Certification Board. n.d. BACB Resources. https://www.bacb.com/about-behavior-analysis/.

Belisle, J., M. Dixon, A. Malkin, J. Hollie, and C. Stanley. 2022. "Exploratory Factor Analysis of the VB-MAPP: Support for the Interdependency of Elementary Verbal Operants." *Journal of Behavioral Education* 31, no. 3: 503–523. https://doi.org/10.1007/s10864-020-09413-2.

Ben-Itzchak, E., L. R. Watson, and D. A. Zachor. 2014. "Cognitive Ability Is Associated With Different Outcome Trajectories in Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 44, no. 9: 2221–2229. https://doi.org/10.1007/s10803-014-2091-05.

Breaux, C. A., and K. Smith. 2023. "Assent in Applied Behaviour Analysis and Positive Behaviour Support: Ethical Considerations and Practical Recommendations." *International Journal of Developmental Disabilities* 69, no. 1: 111–121. https://doi.org/10.1080/20473869.2022.2144969.

Bürkner, P. C. 2017. "BRMS: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80, no. 1: 1–28. https://doi.org/10.18637/jss.v080.i01.

Casula, L., M. G. Logrieco, G. D'Urso, et al. 2024. "What Factors Predict Adaptive Functioning in Preschool Children With Autism Spectrum Disorder? A Longitudinal Study." *Journal of Clinical Medicine* 13, no. 6: 1565. https://doi.org/10.3390/jcm13061565.

Cohen, H., M. Amerine-Dickens, and T. Smith. 2006. "Early Intensive Behavioral Treatment: Replication of the UCLA Model in a Community Setting." Supplement, *Journal of Developmental and Behavioral Pediatrics* 27, no. 2: S145–S155. https://doi.org/10.1097/00004703-200604002-00013.

Cohen, J., P. Cohen, S. G. West, and L. S. Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed. Routledge.

Cooper, J. O., T. E. Heron, and W. L. Heward. 2020. *Applied Behavior Analysis*. 3rd ed. Pearson.

Council of Autism Service Providers. 2024. *Applied Behavior Analysis Practice Guidelines for the Treatment of Autism Spectrum Disorder: Guidance for Healthcare Funders, Regulatory Bodies, Service Providers, and Consumers*. 3rd ed. Council of Autism Service Providers.

Council of Autism Service Providers. 2025. *Evidence About ABA Treatment for Young Children With Autism: The Impact of Treatment Intensity on Outcomes*. Council of Autism Service Providers.

Dixon, M. R. 2014. *PEAK Relational Training System: Direct Training Module*. Shawnee Scientific Press.

Dixon, M. R., J. Belisle, C. Stanley, K. Rowsey, J. Daar, and S. Szekely. 2015. "Toward a Behavior Analysis of Complex Language for Children With Autism: Evaluating the Relationship Between PEAK and the VB-MAPP." *Journal of Developmental and Physical Disabilities* 27, no. 2: 223–233. https://doi.org/10.1007/s10882-014-9410-4.

Eldevik, S., R. P. Hastings, J. C. Hughes, E. Jahr, S. Eikeseth, and S. Cross. 2009. "Meta-Analysis of Early Intensive Behavioral Intervention for Children With Autism." *Journal of Clinical Child and Adolescent Psychology* 38, no. 3: 439–450. https://doi.org/10.1080/15374410902851739.

Eldevik, S., B. Stromgren, S. Eikeseth, A. Fields, C. M. Goetz, and K. B. Titlestad. 2026. "Clinically Significant Outcomes of Early Intensive Behavioral Intervention for Children With Autism Spectrum Disorders:

10990780x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

An Individual Participant Data meta-analysis." *Autism Research* 19: 1–32. https://doi.org/10.1002/aur.70169.

Farley, M. A., W. M. McMahon, E. Fombonne, et al. 2009. "Twenty-Year Outcome for Individuals With Autism and Average or Near-Average Cognitive Abilities." *Autism Research: Official Journal of the International Society for Autism Research* 2, no. 2: 109–118. https://doi.org/10.1002/aur.69.

Frazier, T. W., E. W. Klingemier, C. J. Anderson, G. W. Gengoux, E. A. Youngstrom, and A. Y. Hardan. 2021. "A Longitudinal Study of Language Trajectories and Treatment Outcomes of Early Intensive Behavioral Intervention for Autism." *Journal of Autism and Developmental Disorders* 51, no. 12: 4534–4550. https://doi.org/10.1007/s10803-021-04900-5.

Geiger, K. B., J. E. Carr, L. A. LeBlanc, N. M. Hanney, A. S. Polick, and M. R. Heinicke. 2012. "Teaching Receptive Discriminations to Children With Autism: A Comparison of Traditional and Embedded Discrete-Trial Teaching." *Behavior Analysis in Practice* 5, no. 2: 49–59. https://doi.org/10.1007/BF03391823.

Goldstein, S., and J. A. Naglieri. 2009. *Autism Spectrum Rating Scales (ASRS™)*. Multi-Health System.

Howard, J. S., H. G. Stanislaw, G. Green, C. R. Sparkman, and H. G. Cohen. 2014. "Comparison of Behavior Analytic and Eclectic Early Interventions for Young Children With Autism After Three Years." *Research in Developmental Disabilities* 35, no. 12: 3326–3344. https://doi.org/10.1016/j.ridd.2014.08.021.

Hyman, S. L., S. E. Levy, S. M. Myers, and Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. 2020. "Identification, Evaluation, and Management of Children With Autism Spectrum Disorder." *Pediatrics* 145, no. 1: e20193447. https://doi.org/10.1542/peds.2019-3447.

Jones, S. H., C. C. St. Peter, and M. M. Ruckle. 2020. "Reporting of Demographic Variables in the *Journal of Applied Behavior Analysis*." *Journal of Applied Behavior Analysis* 53, no. 3: 1304–1315. https://doi.org/10.1002/jaba.722.

Kanne, S. M., A. J. Gerber, L. M. Quirmbach, S. S. Sparrow, D. V. Cicchetti, and C. A. Saulnier. 2011. "The Role of Adaptive Behavior in Autism Spectrum Disorders: Implications for Functional Outcome." *Journal of Autism and Developmental Disorders* 41, no. 8: 1007–1018. https://doi.org/10.1007/s10803-010-1126-4.

Klintwall, L., S. Eldevik, and S. Eikeseth. 2015. "Narrowing the Gap: Effects of Intervention on Developmental Trajectories in Autism." *Autism* 19, no. 1: 53–63. https://doi.org/10.1177/1362361313510067.

LeBlanc, L. A., C. Mintz, A. Caffee, and B. A. Kaplan. 2025. "School Placement Outcomes Following Early Intensive Behavioral Intervention in a Routine Clinical Care Setting." *Journal of Autism and Developmental Disorders*: Advance Online Publication. https://doi.org/10.1007/s10803-025-06880-2.

Lord, C., M. Rutter, P. C. DiLavore, and S. Risi. 2012. *Autism Diagnostic Observation Schedule–Second Edition (ADOS-2)*. Western Psychological Services.

Lotfizadeh, A. D., B. Gard, C. Rico, A. Poling, and K. R. Choi. 2025. "Convergent and Discriminant Validity of the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP) and the Vineland Adaptive Behavior Scales (VABS)." *Journal of Autism and Developmental Disorders* 55, no. 3: 803–811. https://doi.org/10.1007/s10803-024-06241-5.

Lotfizadeh, A. D., E. Kazemi, P. Pompa-Craven, and S. Eldevik. 2020. "Moderate Effects of Low-Intensity Behavioral Intervention." *Behavior Modification* 44, no. 1: 92–113. https://doi.org/10.1177/0145445518796204.

Lüdecke, D., M. Ben-Shachar, I. Patil, et al. 2022. "Easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting." CRAN: Contributed Packages. https://doi.org/10.32614/CRAN.package.easystats.

MacDonald, R., D. Parry-Cruwys, S. Dupere, and W. Ahearn. 2014. "Assessing Progress and Outcome of Early Intensive Behavioral Intervention for Toddlers With Autism." *Research in Developmental Disabilities* 35, no. 12: 3632–3644. https://doi.org/10.1016/j.ridd.2014.08.036.

Makowski, D., M. Ben-Shachar, and D. Lüdecke. 2019. "bayestestR: Describing Effects and Their Uncertainty, Existence and Significance Within the Bayesian Framework." *Journal of Open Source Software* 4, no. 40: 1541. https://doi.org/10.21105/joss.01541.

Montallana, K. L., B. M. Gard, A. D. Lotfizadeh, and A. Poling. 2019. "Inter-Rater Agreement for the Milestones and Barriers Assessments of the Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP)." *Journal of Autism and Developmental Disorders* 49, no. 5: 2015–2023. https://doi.org/10.1007/s10803-019-03879-4.

Morey, R., and J. Rouder. 2026. "*BayesFactor: Computation of Bayes* Factors for Common Designs." R package version 0.9.12-4.7. https://doi.org/10.32614/CRAN.package.BayesFactor.

Morey, R. D., R. Hoekstra, J. N. Rouder, M. D. Lee, and E. Wagenmakers. 2016. "The Fallacy of Placing Confidence in Confidence Intervals." *Psychonomic Bulletin & Review* 23, no. 1: 103–123. https://doi.org/10.3758/s13423-015-0947-8.

Myers, A. J., E. Cleveland, P. J. S. Whitby, et al. 2022. "Analysis of a Statewide Early Intervention Program for Young Children With ASD." *Journal of Autism and Developmental Disorders* 52, no. 11: 4994–5006. https://doi.org/10.1007/s10803-021-05376-z.

Padilla, K. L. 2020. "Global Assessment Use and Practices in Applied Behavior Analysis: Surveying the Field." *Research in Autism Spectrum Disorders* 79: 101676. https://doi.org/10.1016/j.rasd.2020.101676.

Padilla, K. L., and J. S. Akers. 2021. "Content Validity Evidence for the Verbal Behavior Milestones Assessment and Placement Program." *Journal of Autism and Developmental Disorders* 51, no. 11: 4054–4066. https://doi.org/10.1007/s10803-020-04864-y.

Padilla, K. L., R. Weston, G. B. Morgan, P. Lively, and N. O'Guinn. 2023. "Validity and Reliability Evidence for Assessments Based in Applied Behavior Analysis: A Systematic Review." *Behavior Modification* 47, no. 1: 247–288. https://doi.org/10.1177/01454455221098151.

Partington, J. W. 2010. *The ABLLS-R—The Assessment of Basic Language and Learning Skills–Revised*. Behavior Analysts Inc.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Reichow, B. 2012. "Overview of Meta-Analyses on Early Intensive Behavioral Intervention for Young Children With Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 42, no. 2: 512–520. https://doi.org/10.1007/s10803-011-1218-9.

Reichow, B., K. Hume, E. E. Barton, and B. A. Boyd. 2018. "Early Intensive Behavioral Intervention (EIBI) for Young Children With Autism Spectrum Disorders (ASD)." *Cochrane Database of Systematic Reviews* 5, no. 5: CD009260. https://doi.org/10.1002/14651858.CD009260.pub3.

Reichow, B., and M. Wolery. 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, no. 1: 23–41. https://doi.org/10.1007/s10803-008-0596-0.

Ridout, S., and S. Eldevik. 2024. "Measures Used to Assess Treatment Outcomes in Children With Autism Receiving Early and Intensive Behavioral Interventions: A Review." *Review Journal of Autism and Developmental Disorders* 11, no. 3: 607–619. https://doi.org/10.1007/s40489-023-00355-9.

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Schopler, E., M. E. Van Bourgondien, G. J. Wellman, and S. R. Love. 2010. *Childhood Autism Rating Scale–2nd Edition*. Western Psychological Services.

Shaw, K. A., S. Williams, M. E. Patrick, et al. 2025. "Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years—Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022." *MMWR Surveillance Summaries* 74, no. SS-2: 1–22. http://dx.doi.org/10.15585/mmwr.ss7402a1.

Skinner, B. F. 1957. *Verbal Behavior*. Appleton.

Smith, T., R. Klorman, and D. W. Mruzek. 2015. "Predicting Outcome of Community-Based Early Intensive Behavioral Intervention for Children With Autism." *Journal of Abnormal Child Psychology* 43, no. 7: 1271–1282. https://doi.org/10.1007/s10802-015-0002-2.

Sparrow, S. S., D. V. Cicchetti, and D. A. Balla. 2005. *Vineland Adaptive Behavior Scales: Second Edition (Vineland II)*. Pearson Assessments.

Sparrow, S. S., D. V. Cicchetti, and C. A. Saulnier. 2016. *Vineland Adaptive Behavior Scales: Third Edition (Vineland-3)*. NCS Pearson.

Sundberg, M. L. 2014. *The Verbal Behavior Milestones Assessment and Placement Program: The VB-MAPP*. 2nd ed. AVB Press.

Szatmari, P., S. Georgiades, E. Duku, et al. 2015. "Developmental Trajectories of Symptom Severity and Adaptive Functioning in an Inception Cohort of Preschool Children With Autism Spectrum Disorder." *JAMA Psychiatry* 72, no. 3: 276–283. https://doi.org/10.1001/jamapsychiatry.2014.2463.

Vietze, P., and L. Lax. 2020. "Early Intervention ABA for Toddlers With ASD: Effect of Age and Amount." *Current Psychology* 39, no. 4: 1234–1244. https://doi.org/10.1007/s12144-018-9812-z.

Virués-Ortega, J. 2010. "Applied Behavior Analytic Intervention for Autism in Early Childhood: Meta-Analysis, Meta-Regression, and Dose-Response Meta-Analysis of Multiple Outcomes." *Clinical Psychology Review* 30, no. 4: 387–399. https://doi.org/10.1016/j.cpr.2010.01.008.

Volkmar, F. R., R. Paul, S. J. Rogers, and K. A. Pelphrey. 2014. *Handbook of Autism and Pervasive Developmental Disorders*. 4th ed. John Wiley & Sons.

Waters, C. F., M. Amerine Dickens, S. W. Thurston, X. Lu, and T. Smith. 2020. "Sustainability of Early Intensive Behavioral Intervention for Children With Autism Spectrum Disorder in a Community Setting." *Behavior Modification* 44, no. 1: 3–26. https://doi.org/10.1177/0145445518786463.

Weisz, J. R., A. M. Ugueto, D. M. Cheron, and J. Herren. 2013. "Evidence-Based Youth Psychotherapy in the Mental Health Ecosystem." *Journal of Clinical Child and Adolescent Psychology* 42, no. 2: 274–286. https://doi.org/10.1080/15374416.2013.764824.

Wergeland, G. J. H., M. Posserud, K. Fjermestad, U. Njardvik, and L. Öst. 2022. "Early Behavioral Interventions for Children and Adolescents With Autism Spectrum Disorder in Routine Clinical Care: A Systematic Review and Meta-Analysis." *Clinical Psychology: Science and Practice* 29, no. 4: 400–414. https://doi.org/10.1037/cps0000106.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag: ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org.

World Health Organization. 2016. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. World Health Organization.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information S1: bin70142-sup-0001-suppl-data.docx.

1099078x, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bin.70142 by Linda LeBlanc - Behavior Analyst Certification Board , Wiley Online Library on [11/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License