ASPECTS OF INTENSIVE BEHAVIOURAL INTERVENTION QUALITY AND THEIR RELATIONSHIP WITH CHILD CHARACTERISTICS AND OUTCOMES

KSUSHA OKSANA BLACKLOCK

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

May 2016

Abstract

Intensive Behavioural Intervention (IBI) is currently the treatment of choice for young children with moderate to severe Autism Spectrum Disorders. Research has examined different aspects of the intervention, though there is a paucity of information on the quality of IBI. This study examined the York Measure of Quality of IBI (YMQI) in 39 children receiving publicly-funded IBI in Ontario for approximately one year. Videos (*n* = 402) of children engaged in IBI were coded using the YMQI. A factor analysis showed that the YMQI measures different aspects of IBI quality: Pace and Organization, Technical Correctness, Engagement and Motivation, and Generalization. All of these subscales remained fairly stable, within the "good" quality range, over approximately one year in IBI, with relatively lower Generalization scores. An examination of the relationships between the YMQI subscales and children's characteristics at the start of IBI showed that children with more autism symptomatology at the start of treatment receive intervention lower in Engagement and Motivation at the beginning of treatment. In terms of the connection of IBI quality to children's progress, there was a relationship between Technical Correctness at the start of treatment and greater decreases in autism severity, as well as relationships between Generalization and children's gains in cognitive skills and decreases in autism severity. Finally, the change in autism symptomatology was predicted by three of the trajectories of quality subscales throughout the year. These results help operationalize the quality of IBI more precisely and have implications for IBI training, supervision, and research.

Keywords: Autism Spectrum Disorder (ASD), Intensive Behavioural Intervention (IBI), treatment quality.

Acknowledgments

This research would not have been possible without the help and support of many others. First and foremost, I am forever grateful to Dr. Adrienne Perry. Not only has she been most instrumental in the completion of this research but also in supporting me throughout the graduate school and dissertation process. I cannot imagine completing this project without her continuous support, understanding, and encouragement. Her commitment and confidence in me, as well as her helpful and patient guidance, made me feel supported throughout the entire project and I cannot thank her enough for that. I realize how very lucky I am to have her as my supervisor and am forever grateful for the opportunity to have worked and grown with her during this project.

I would like to extend my sincere gratitude to Dr. David Flora for agreeing to be on my committee and spending innumerable hours helping me to narrow down my research questions, to run the complicated statistical analyses, and to interpret my results in a meaningful and applicable way. His honest feedback and revisions to this project were essential in its completion.

I would also like to thank Dr. Nancy Freeman for being an integral part of this research project from the beginnings of ethics applications and data collection, to its completion. Her quick responses to my questions and insightful edits to the manuscript have allowed me to see this research from a different and more applied perspective and made this research much more applicable to the real world.

An enormous thank you goes out to my lab mates, and friends, who make the Perrylab such a wonderful and supportive environment. Their advice and perspectives throughout the years have been extremely helpful in keeping the end goal in mind and in continuing to persevere through the more challenging parts of the process.

This research would not have been possible without the participation of the families who agreed to take part in this project. In addition, a big thank you goes out to the Children's Hospital of Eastern Ontario, Kinark Child and Family Services, and Surrey Place Centre for allowing me to conduct this research in your centres. Thank you for opening your doors and seeing the value of this project.

Finally, I would like to thank my family for their support throughout this journey, in particular, my husband, Sean Blacklock. B, thank you for believing in me, encouraging me every step of the way, and for going through all the highs and lows of this journey with me. Your love and support has been vital to the completion of this research and I could not have done it without you by my side. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Aspects of Intensive Behavioural Intervention Quality and their Relations to Child Characteristics and Outcomes

Autism

Autism Spectrum Disorder (ASD) is a neurodevelopmental and behaviourally defined disorder characterized by persistent impairment in reciprocal social communication and social interaction, and restricted, repetitive patterns of behaviour, interests or activities, with onset early in life (American Psychiatric Association, 2013). The latest studies in North America, Europe, and Asia estimate the average prevalence of ASD in the population to be about 1% (Malcolm-Smith, Hoogenhout, Ing, Thomas, & de Vries, 2013), which reflects a significant increase in recent years (Elsabbagh, et al., 2012). Across these and other studies, great variability is found in children's ASD presentation as well as their response to intervention (Eldevik et al., 2009).

Intensive Behavioural Intervention

Research indicates that Intensive Behavioural Intervention (IBI) is currently the treatment of choice for young children falling toward the moderate to severe end of the Autism Spectrum (Peters-Scheffer, Didden, Korzilius & Sturmey, 2011; Reichow, 2012). IBI is based on the principles and strategies of Applied Behaviour Analysis (ABA) and involves individual intervention for a substantial number of hours per week, targeting differences between the skills of children who have developmental delays and the skills of their typically developing peers (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005). Studies have shown better outcomes for children participating in IBI than in other eclectic and community-based programs (Eikeseth, Smith, Jahr, & Eldevik, 2007), special education classes in public schools (Cohen, Amerine-Dickens, & Smith, 2006), eclectic “treatment-as-usual” programs which incorporate elements from a variety of interventions such as ABA, sensory-motor therapies, and Treatment and Education of Autistic and Related Communication Handicapped Children (Eikeseth, Klintwall, Jahr, & Karlsson, 2012; Eldevik, Hastings, Jahr, & Hughes, 2012; Howard et al., 2005), even when these interventions have equal treatment hours to IBI programs (Eikeseth, Smith, Jahr, & Eldevik, 2002; 2007; Howard et al., 2005).

IBI has been shown to improve the skills of children with ASD in several domains, such as cognitive functioning (Lovaas, 1987; Perry, Cummings, et al., 2008; Sheinkopf & Siegel, 1998; Smith, Groen, & Wynn, 2000), adaptive behaviour (Magiati, Charman, & Howlin, 2007; Remington et al., 2007; Sallows & Graupner, 2005), and language skills (Hayward, Eikeseth, Gale, & Morgan, 2009; Howard et al., 2005). Eldevik and colleagues (2009) reported a meta-analysis of nine controlled studies of IBI and found a large standardized mean difference effect size for IQ (1.10) and a medium effect size for adaptive behaviour (.66). For example, Flanagan, Perry, and Freeman (2012) found that the children who took part in IBI showed better outcomes than a matched waitlist group in all outcomes measured, including autism severity, adaptive functioning, and cognitive skills.

Despite this promising body of literature, IBI has not worked for all children (Howlin, Magiati, & Charman, 2009) as studies have found that some children demonstrate little to no progress following IBI (e.g., Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Birnbrauer & Leach, 1993), and some even demonstrate fewer skills following a period of IBI (Eldevik, Eikeseth, Jahr, & Smith, 2006; Perry, Cummings, et al., 2008; Remington et al., 2007). In their meta-analysis of IBI effectiveness, Reichow and Wolery (2009) reviewed available individual data for each participant and found that at least one participant receiving IBI in each study showed no improvement or lost skills on at least one outcome variable. This heterogeneity in outcomes may be explained by various factors that have been suggested to affect a child's outcome in IBI, including child, family, and treatment variables (Perry & Freeman, 1996; Perry, Koudys, & Blacklock, 2016), described further in the next sections.

Child Characteristics Affecting IBI Outcomes

Research has been conducted on child factors that may affect treatment outcome in IBI. In particular, cognitive level (IQ) and age at start of treatment have received much attention. Many researchers (Ben-Itzchak, Watson, & Zachor, 2014; Eikeseth et al., 2002; 2007; Harris & Handleman, 2000; Hayward et al., 2009; Sallows & Graupner, 2005; Smith, Klorman, & Mruzek, 2015) have found that a child's IQ at the start of treatment is moderately or highly correlated with the child's outcome in IBI, while Remington et al. (2007) showed that initial IQ predicted which children were most responsive to treatment. Beglinger and Smith (2005), Perry et al. (2011), and Magiati et al. (2007) found that initial IQ accounted for large portions of the variance in outcomes following IBI (32%, 54%, and 35%, respectively). In contrast, a few studies have found initial IQ not to be significantly related to outcome (Cohen et al., 2006; Smith et al., 2000). In addition, Eikeseth and colleagues (2007) demonstrated that higher IQ is correlated with better outcomes for children regardless of whether they are receiving behavioural treatment or eclectic treatment.

Several studies have also focused on the effect of a child's age at start of treatment on their intervention outcomes. Many studies have shown that a younger starting age was associated with improved outcomes (Bibby, Eikeseth, Martin, Mudford, & Reeves, 2002; Eldevik et al., 2012; Fenske, Zalenski, Krantz, & McClannahan, 1985; Flanagan et al., 2012; Freeman & Perry, 2010; Goin-Kochel, Myers, Hendricks, Carr, & Wiley, 2007; Perry, Blacklock, & Dunn Geier, 2013; Perry et al., 2011; Smith et al., 2015). Recently, Perry et al. (2013) showed that, even after controlling for initial IQ, younger age was predictive of better cognitive outcomes including IQ after treatment, cognitive rate of development during IBI, and especially change in IQ.

There are also studies that have shown that age at start of treatment may not be a significant predictor of children's outcomes in IBI (Eikeseth et al., 2002; 2007; Eldevik et al. 2006; Hayward et al., 2009; Klintwall, Eldevik, & Eikeseth, 2015; Lovaas, 1987; Magiati et al., 2007; Sallows & Graupner, 2005; Sheinkopf & Siegel, 1998). However, as Perry and colleagues (2011) noted, small sample sizes (e.g., 11 children) and limited age ranges (e.g., 24 to 42 months in Hayward et al., 2009; 16 to 46 months in Lovaas, 1987) likely prevented these studies from detecting significant effects of age, as all participants fell within the optimal age range for intervention. In addition, a more recent study found that, even among children aged 1 to 3 years at the start of treatment, it was the youngest children, those who began intervention between the ages of 18 and 23 months, that improved the most (MacDonald, Parry-Cruwys, Dupere, & Ahearn, 2014).

The various meta-analyses examining child characteristics report conflicting results regarding the effects of pre-treatment age and IQ on children's outcomes in IBI, with one finding that age was important (Makrygianni & Reed, 2010), another concluding that IQ was important while age was not (Howlin et al., 2009), and yet another finding that neither pre-treatment age nor IQ affected outcomes (Reichow &

Wolery, 2009). These conclusions are clearly limited to the particular sample characteristics, e.g., the age range and IQ range, included in the studies.

Treatment Characteristics Affecting IBI Outcomes

The effect of intervention factors on children's variable outcomes in IBI has also been explored. Generally, higher "intensity" of intervention (20-40 hours per week) has been linked to improved outcomes for children (Eikeseth et al., 2002; Granpeesheh, Kenzer, & Tarbox, 2011; Lovaas, 1987; Smith et al., 2000). Makrygianni and Reed (2010) conducted a meta-analysis of 14 studies and found that the intensity (number of hours per week) and duration of the programs were correlated with program effectiveness. Virues-Ortega, Rodriguez, and Yu (2013) found that total intervention time in hours, which is a combination of both treatment intensity and treatment duration, was the single predictor with the highest contribution, regardless of pre-intervention functioning or age.

IBI Quality

Though quantity of intervention is clearly important, what happens during all those hours is also undoubtedly important to children's outcomes (Bibby et al., 2002; Green, 1996; Perry, 2002). This is referred to as treatment quality, and has rarely been examined in the IBI literature. There is no current consensus on the measurement of quality in IBI, although some have examined treatment fidelity in therapists providing home-based behavioural intervention (Symes, Remington, Brown, & Hastings, 2006), or adherence in a parent-mediated behavioural program (Allen & Warzak, 2000). Klintwall, Gillberg, Bölte, and Fernell (2012) examined how therapist allegiance to IBI and ABA techniques impacts children's treatment outcomes, represented by change in adaptive behaviour scores over the duration of the treatment, and found that therapist allegiance accounted for 19% of the variance in children's outcomes, supporting the idea that allegiance to behavioural techniques mediates the effectiveness of treatment. The supervisor training model, which may be seen as a mechanism to ensure quality, was associated with better outcomes in Reichow and Wolery's (2009) meta-analysis. There is also some evidence to suggest that programs which are of uncertain quality (e.g., low number of hours, poor training, and poor supervision) are associated with less optimal outcomes (e.g., Bibby et al., 2002).

Despite these efforts, quality of IBI remains a difficult concept to measure and study and has not been examined in large, real-world IBI programs. With such variable outcomes and the large amount of public money being spent on programs such as the Ontario provincial IBI program, measuring quality is extremely important. Over a decade ago, Jacobson (2000) raised concerns about quality-control in early intensive intervention echoing concerns of earlier professionals in behaviour analysis (Johnston & Sherman, 1993; Shook, 1993; Wood, 1975). At the time he suggested that "...quality-control problems represent profound threats to the effectiveness of intensive early intervention services and policy, and pose the very real possibility that, unless action is taken to correct them, intensive early intervention policy will become a debacle and eventually be looked back upon as a travesty of human and public service" (pp. 162-163). To address this issue, professionals in behaviour analysis instituted a certification process, which was intended to help address adequate preparation of personnel in behavioural methods and ensure that people and organizations that provide behavioural intervention are competent (Shook, Hartsfield, & Hemingway, 1995).

While the above-mentioned efforts to improve quality control from the perspective of staff training and credentialing are important, it is also imperative to examine the quality of the intervention actually received by children. An initial delineation of the important ingredients of IBI quality was made by Perry, Prichard, and Penn (2006). They asked parents and professionals to choose important characteristics of high quality IBI from a list of 11 generally accepted characteristics of quality teaching gleaned from the literature, staff training manuals, and clinical experience and they also asked participants how these characteristics should be measured. Fifty-two professionals, including clinical directors, supervising psychologists, supervisors, and senior therapists, completed surveys. As important aspects of quality teaching, clinical directors most often chose administering reinforcers of the appropriate type, creating opportunities for generalization, setting up opportunities for child directed learning, and using effective and appropriate behaviour management strategies. Supervisors also endorsed administering reinforcers of the appropriate type and creating opportunities for generalization but, unlike other groups, emphasized administering prompts of the appropriate type as a significant indicator of quality teaching. The group of 'other professionals' responded in a similar pattern, emphasizing the importance of creating opportunities for generalization, using effective and appropriate behaviour management strategies, administering reinforcers of the appropriate type, and at the appropriate time. Overall, professionals also indicated a strong preference for measuring these characteristics objectively versus subjectively.

The survey was also completed by 20 parents of children with ASD involved in their child's IBI programs (e.g., using behavioural principles, attending team meetings, setting goals, observing sessions), including seven parents who reported conducting intensive teaching sessions with their child(ren). In general, parents rated most of the characteristics as very important or important, with the notable exception of recording data on children's trial-by-trial performance. Parents thought that varying task presentation, creating opportunities for generalization, and using effective and appropriate behaviour management strategies were most important to quality teaching, emphasizing the importance of task variation.

Comparing the answers of parents and clinical directors revealed that parents stressed the importance of varying the discriminative stimuli and therapist characteristics (e.g., enthusiasm), whereas clinical directors stressed the more technical aspects of IBI such as reinforcement type, program design, and supervision and training. Combining both groups, the three most frequently endorsed characteristics of quality IBI teaching were creating opportunities for generalization (endorsed by 49% of participants), administering reinforcers of the appropriate type (44%), and using effective and appropriate behaviour management strategies (38%). However, none of the characteristics was universally endorsed.

Both versions of the survey contained an open-ended question about additional characteristics of high quality IBI, and participants suggested many other important characteristics of high quality IBI teaching other than the ones provided. These responses were grouped thematically. The category containing the most comments captured information related to program design (the program is well-matched to the child's skill level, the program targets a wide range of functional skills, the program incorporates a variety of teaching strategies and settings, etc.). The four other categories derived from participants' suggestions were: the program is linked to a larger context such as home and school (36 comments); the therapist has appropriate skills and experience (29 comments); the program involves regular supervision and training (28 comments); and the program is applied and adapted appropriately (18 comments).

History and Development of the YMQI

Based on these results, Perry, Flanagan, and Prichard (2008) set out to create a valid and reliable measure of IBI quality. The York Measure of Quality of IBI (YMQI) was created based on accepted practices for the development of observational measures, such as outlining behavioural categories, carrying out pilot observations, creating operational definitions, determining response dimensions, and outlining the measurement context (Hartmann & Wood, 1990). Behavioural categories relevant to quality teaching were gleaned from several sources including the survey results discussed above, training manuals published by expert clinicians (e.g., Lovaas, 2003), experimental research suggesting that specific teaching procedures facilitate learning and generalization (e.g., Stokes & Baer, 1977), rating scales used by treatment providers to monitor staff performance (Hundert, Walton-Allen, Earle-Williams, Sim, & Cope-Scott, 2000; Leaf, & McEachin, 1999; Provincial Regional Trainers Network, 2004), and the only two empirically supported measures in the literature which were designed to evaluate staff competence within prescribed teaching situations using a specific type of IBI (Davis, Smith, & Donahoe, 2002; Koegel, Russo, & Rincover, 1977). Throughout the process, expert psychologists in IBI were consulted to incorporate clinical judgment into the measure.

Following the literature review and the survey, the authors designed the Pilot Version of the YMQI and assessed its psychometric properties. The Pilot Version included 30 items in nine categories: Discriminative Stimuli (S^Ds), Reinforcement, Prompting, Learning, Pacing, Engagement, Generalization, Problem Behaviour, and Organization. Each item was scored according to an "objective" scale using 30-second interval coding and a "subjective" scale, a rating approach with 5-point Likert-type response options. Three 3-minute segments of videotapes of 28 therapist-child dyads engaged in IBI with local public and private service providers were coded by six raters trained to criterion. In addition, four experts in IBI who were supervising psychologists or Board Certified Behavior Analysts provided expert judgments to facilitate evaluations of criterion-related validity.

Prichard (2005) evaluated the reliability of this version. She found that the internal consistency was good and that inter-rater reliability (IRR) varied from poor to good on different items of the two scales. Penn (2005) examined the validity of the measure and reported that content validity was strong; construct validity, based on the inter-relationships of the items, was good; and criterion-related validity, based on the relationship to expert judgment, was moderate to good on different items and categories on the two scales. Overall, the results suggested that the "subjective" approach (Likert-type ratings) to evaluating the quality of IBI was preferable to the objective approach (30-second interval coding), as the subjective ratings were more highly correlated with expert judgment, more internally consistent, possessed better IRR across items and categories, and took substantially less time to complete (1 hour versus 6 hours). There were several items with weaker psychometric properties, suggesting that refinements to the operational definitions or coding rules were needed. Penn, Prichard, and Perry (2007) concluded that changes were necessary prior to the measure's implementation.

For subsequent versions, the authors focused on the subjective scale, but aimed to make it as objective as possible by specifying clear operational definitions and explicit rules about how to rate each item. The YMQI was revised in 2006 and its psychometric properties re-evaluated (Prichard, Penn, Perry, Solish, & Levy, 2007). This version of the YMQI had 33 items in nine categories: S^Ds, Reinforcement, Prompting, Organization, Pacing, Teaching Level, Instructional Control, Generalization, and Problem Behaviour. Four new raters were trained to criterion and rated 36 videotapes (including some new tapes and new segments from some of the tapes obtained for the earlier study). Four expert raters, psychologists and/or behaviour analysts, provided the expert judgments. The reliability and validity of this 2006 version of the YMQI was reported by Prichard et al. (2007). Internal consistency was excellent ($\alpha = .86$) with moderate to strong item-total correlations for 21 of the items. IRR (using intraclass correlations [ICC]) was acceptable for 26 items, six categories, and the total score (mean of individual items ICC = .61). Criterion-related validity, relative to expert judgment, was adequate for 28 items, five categories, and the YMQI total score ($r = .58$ with Expert Judgment Scale). Based on these results, two items (Wording of S^Ds, Speed of prompting) were dropped or combined into existing items and the reliance on categories was de-emphasized.

The Current Version of the YMQI

The current version of the YMQI includes 31 items in nine categories. It involves coding two 5-minute segments of videotaped IBI sessions and can be applied to any IBI instructional methodology or curriculum goal area, including discrete trial teaching, natural environment teaching, analysis of verbal behaviour, and fluency-based approaches.

Because the YMQI is a measure with established reliability and validity, proper training is required to implement it accurately. The training includes a manual with detailed coding rules for each item, a behavioural principles quiz, a self-guided training DVD and achieving inter-rater agreement with the YMQI developers on training videos.

In order to use the YMQI, a video of an IBI session at least 20 minutes in length must be obtained and two 5-minute segments chosen at random. The coders are then required to watch each 5-minute segment and code it according to the detailed instructions outlined in the YMQI Administration Manual (Perry, Flanagan, et al., 2008). Each segment is scored according to 31 individual items in nine categories, each containing two to six items (see Table 1).

There are two different types of items: frequency items and evidence items. Frequency items (items 1, 3, 4, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 29) are rated according to the frequency of correct teaching skills observed during the segment. In other words, coders rate based on the proportion of time that the therapist is behaving according to the "Good Practice" guidelines outlined for each item, while also looking for various "Mistakes" or problems that may occur, including whether any "Significant Mistakes" or more serious problems were present in the segment. Some items simply require the coder to note how many mistakes and significant mistakes were made and rate the item according to its criteria (items 3, 4, 9, 11, 13, 19, and 29), while other items also require the coder to note how much of the total time (e.g., total number of trials, reinforcers, prompts, etc.) the therapist was correct in her administration of the item (items 1, 7, 12, 14, 15, 16, 17, 21, and 22). For example, item '3. Rapid reinforcer delivery' rates whether tangible or activity reinforcers are delivered quickly enough for the child to link them with the correct response. Good practice for this item is outlined as the therapist having the reinforcer ready when the child responds and delivering it quickly, contingent on the child's correct response. The good practice description goes into detail and outlines that there should be no more than 2 seconds between the end of the child's response and the time that the reinforcer begins. Mistakes for this item occur when a reinforcer is delivered more than 3 seconds after the child's response. A significant mistake occurs when the time between the child's response and the reinforcer is very long (approximately 10 seconds), or when the time between the child's response and the reinforcer is shorter, but the child engages in a different behaviour before the reinforcer is received, such that the wrong behaviour may be reinforced. The item should receive a rating of 3 (very good) if there were no mistakes in the segment, a rating of 2 (generally good) if two or three mistakes but no significant mistakes were observed in the segment, and 1 (poor) if more than one significant mistake occurred.

The second type of items, Evidence items (2, 5, 6, 8, 10, 20, 23, 24, 25, 26, 27, 28, 30, 31), are rated according to pre-established evidence criteria. These items reflect higher scores for evidence of particular aspects of good teaching and do not have any clearly defined mistakes to watch for. In contrast, there are ways in which therapists can display "Evidence" of high quality teaching and situations in which they could have shown evidence of high quality teaching but did not, referred to as "Missed Opportunities". For example, item '2. Varying S^Ds' rates whether the child is being taught using a variety of instructions, such as slightly different wording and tone of voice in the therapist's instructions to the child, even when the therapist is asking for the same response. There is "evidence" for this criterion when the therapist varies the presentation of an exemplar two or more times during the segment, and there is a "missed opportunity" when there are many repeated trials for the same exemplar (five or more) for which the therapist does not vary the S^D. The item is rated as very good (3) when there is evidence and no missed opportunity, generally good (2) when there is evidence and one or more missed opportunity or no evidence and no missed opportunity, and poor (1) when there is no evidence and one or more missed opportunity.

Both types of items are rated using a 5-point scale, ranging from 1 to 3 with half-points (1, 1.5, 2, 2.5, and 3). The half-points are assigned when the item meets intermediate criteria between a 1 and a 2 or between a 2 and a 3. For some of the items an "N/A" option is available. For example, Item '29. Result of problem behaviour' would be rated "N/A", if no problem behaviour occurred during that segment. A total score is calculated for each segment and each session, which is based on the scores obtained from the two segments.

The psychometric properties of this current version are presented in the Background and Development of the YMQI document in the YMQI package (Perry, Flanagan, et al., 2008). The overall internal consistency for this final version is $\alpha = .82$. The overall IRR, based on averaging 31 items, is ICC = .68, which is considered good. The overall validity coefficient for the total score, based on correlations with expert judgment, is .58, which is considered strong.

The YMQI is one component of a broader York System of Quality Assurance for IBI (YSQA), which includes components that review the quality of teaching, quality of programming, and quality of the organization, along with the YMQI. However, the YMQI may be used alone, without the other components for research or clinical purposes, and is the focus of the current study.

Prospective Study

In 2000, Ontario launched a province-wide IBI initiative based on research and stakeholder consultation (MCSS, 2000; Perry, 2002), in which IBI is funded by the provincial Ministry of Children and Youth Services and is provided free of charge to families. Currently, there are nine Regional Programs with public and private partners and subcontractors, serving approximately 2000 children and costing over $115 million annually (Office of the Auditor General of Ontario, 2013). (For more information about Ontario's IBI program and its effectiveness, see Perry, Cummings, et al., 2008.)

Recently, the YMQI was used in a prospective study in Ontario with the goal of evaluating the ecological validity of the IBI and furthering the understanding of which child, family, and treatment factors, including quality of IBI, can predict children's outcomes (Dunn Geier, Perry, & Freeman, 2012). The study involved assessing children enrolled in Ontario regional IBI programs and comparing them to a wait-list control group over a one-year period. The prospective waitlist-controlled study is the strongest research design possible for IBI effectiveness research, especially since the children in the IBI and waitlist groups are similar in that they are both deemed eligible for the program and parents in both groups want their children to take part in IBI. The primary outcome measures were cognitive level, adaptive behaviour level, and autism severity, each of which was also assessed at the beginning of the study. Additional measures included family variables such as stress, coping, and family demographics. Families were also asked about other services their children may have been receiving before or during IBI or while on the waitlist. The same procedure was repeated one year later. In the IBI group, the amount of treatment and supervision was monitored weekly and parent involvement was also measured. Finally, and of most relevance to the present study, children involved in the IBI group were videotaped once per month during their IBI sessions and these tapes were coded using the YMQI.

The YMQI portion of the prospective study required research assistants or IBI staff to videotape children taking part in IBI for approximately 30 minutes per month for the 12 months of the study (creating 402 videos in total), during regular IBI program activities. Once the coders received the videos, two 5-minute segments were selected as per the YMQI Administration Manual instructions (Perry, Flanagan, et al., 2008). The coders were four undergraduate students who trained to criterion using the YMQI training DVD as well as an in-person training session with the project coordinator (the present author).

Using a sample from this prospective study data set, Whiteford, Blacklock, and Perry (2012) examined the IRR of the current version of the YMQI and the current method of training the coders using 33 videos (25% of the videos coded to that point in the study). IRR was calculated using percentage agreement, which was defined as two coders rating an item within one ½ point of one another. An overall score of at least 80% is considered acceptable for this type of measurement scale (Cooper, Heron, & Heward, 2007). IRR was calculated for individual items and ranged from 74 to 100%, as well as the total score of the YMQI, which ranged from 74 to 97%.

After double-coding 20% of the videos from the prospective study (83 randomly chosen videos), Blacklock, Shine, and Perry (2013) reexamined the IRR and internal consistency of the YMQI. IRR was calculated using both percent agreement (within ½ point) and ICC, one of the most commonly-used statistics for assessing IRR for ordinal variables (Hallgren, 2012). Percent agreement ranged from 76% to 100% (*M* = 88.23, *SD* = 6.80), which is comparable to the partial sample analyzed by Whiteford et al. (2012). The overall 88% agreement is better than was previously reported by Prichard (77%; 2005), which indicates that revisions to the YMQI have improved the IRR. ICCs ranged from -.11 to .85 (*M* = .33, *SD* = .35), with only one third of the items above an ICC of .50. This suggests that items with poor IRR may warrant reexamination, clarification, or being dropped, and, further, suggests that the measure may not be unidimensional.

Often in measures such as the YMQI, an observer's level of accuracy might shift in the months following training; hence potential observer drift over time was also examined. The level of IRR was consistent over time and showed only a very weak correlation with time since training (*r* = -.014) indicating that there was no observer drift over a 9-month period (Whiteford et al., 2012).

Internal consistency, as measured by coefficient alpha, was α = .77, which is acceptable for such a complicated and subjective scale; however, it is somewhat lower than the α = .86 reported by Prichard (2005). Exploration of α coefficients with individual items deleted revealed no particular problematic items. Note that item '29. Result of problem behaviour' was often rated "N/A" and was, therefore, removed from these analyses. Item-total correlations for the remaining 30 individual items ranged from poor to good (-.13 to .55), with 67% (*n* = 20) above .30. Items 4, 'Motivating reinforcers', 5, 'Varying reinforcers', 27, 'Response generalization', and 28, 'Flexible teaching' were poorly correlated with total quality. Whiteford et al. (2012) noted that it might be necessary to reevaluate these items to see if they can be improved, perhaps by rewording their instructions or rating criteria, or possibly drop them from the measure. Even removing both of the two poorest items (item 5, 'Varying reinforcers', and item 28, 'Flexible teaching') would only improve α to .79 (Blacklock, Perry, & Whiteford, 2011). The differences between these and Prichard's (2005) results may be due to revisions to the YMQI, the different coders, the method of training the coders, and the particular sample of videos.

Following these findings, Taheri, Blacklock and Perry (2013) examined whether treatment quality, as measured by the YMQI, is consistent over time, using the data from the prospective study described above. For this study, 15 children were selected who had a video within one month of the IBI start, another video 5 to 7 months later, and a final video after approximately one year in IBI (11 to 13 months after the child began treatment). Three rationally-derived subscales of the YMQI were created, Technical Correctness, Promoting Generalization, and Problem Behaviour, based on theoretical considerations and inter-item correlations. The level and trend of the three subscales over three time points were examined using profile analysis (Tabachnick & Fidell, 2007). No statistically significant trend was detected, indicating good consistency over time. Overall, quality of IBI started out in the "good" range in the first month of treatment and remained so for the other two time points during one year of treatment. When level was examined, Promoting Generalization scores were consistently lower at each time point (most differences were statistically significant), which may be a function of the particular items in this section of the scale, most of which are "evidence" items.

Overall, this research with the YMQI shows that it is a reasonably psychometrically sound measure of the quality of IBI. However, our most recent research also points to the possibility that the YMQI does not measure one unitary construct of the quality of IBI. It is important to know whether the YMQI measures different dimensions or factors of IBI quality, and whether only looking at specific subscales, instead of the total score, may be the most helpful and valid use of the YMQI.

Another important question stems from the fact that initial prospective study analyses showed a significant correlation between the YMQI total score and children’s rate of development during IBI, a finding that needs to be examined in more detail (Dunn Geier et al., 2012). As described earlier, meta-analyses have found that some aspects of IBI are important to children’s outcomes, such as the supervision model, intensity, and duration; however, more research is needed to fully understand the relationship of child and treatment factors and their effect on outcomes. Since few studies have included measures of quality of IBI, many unanswered questions remain regarding this important aspect of the intervention. The present study offers a unique opportunity to examine treatment quality in detail, and to explore its relationships with child characteristics and children's progress in treatment.

Based on a large sample of videos from the prospective study, the current study addresses two sets of research questions: the first focusing on the factor structure of the measure itself; and the second examining treatment quality over time, as well as the relationships of treatment quality to child characteristics and progress during IBI.

Methods

The prospective study was approved by the Research Ethics Board at York University as well as the three participating agencies. The current study was also approved by the Research Ethics Board at York University. The data for the present study come from the children in the IBI group in the previously described study, which had a prospective, quasi-experimental design and included data from three participating sites (the Toronto Partnership for Autism Services, Surrey Place Centre; the Autism Intervention Program – Eastern Ontario, CHEO; and the Central East Autism Program, Kinark Child and Family Services).

Participants

All participants were children who took part in Ontario's publicly funded IBI program. The sample for analysis involved 39 children (30 from Toronto; 5 from Central East; 4 from the Eastern Region). Pre-intervention data (Time 1) were collected within two months before the child began IBI, with some relevant data accessed from children's clinical files (e.g., diagnosis). A summary of participant characteristics at Time 1 is presented in Table 2. Participants were mostly males with a diagnosis of Autistic Disorder, with significantly below average cognitive and adaptive skills. A second assessment (Time 2) was conducted approximately one year after the initial assessment, plus or minus 6 weeks. The present author took part in many of these assessments. All assessments were carried out under the supervision of an experienced registered psychologist. Between the two assessments, children took part in IBI at one of the three participating centres for approximately 20 hours per week.

Measures

Autism severity was assessed using the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988), a behavioural observation measure based on observations made through direct interaction with the child conducted during the course of a psychological assessment. Ratings are supplemented with a parent report for items that cannot be observed during the assessment. A higher total score indicates greater autism severity. Scores fall within three classifications: severe autism, mild/moderate autism, or not autism. The CARS is very reliable, displaying good internal consistency, acceptable inter-rater agreement, and good test-retest stability (Perry, Condillac, Freeman, Dunn Geier, & Belair, 2005; Schopler, et al., 1988).

Cognitive level was most often obtained using the Mullen Scales of Early Learning, while the Wechsler Preschool and Primary Scale of Intelligence Third Edition was used in three cases.

Mullen Scales of Early Learning (Mullen, 1995) is a standardized, norm-referenced measure of children’s level of cognitive functioning, which produces four “cognitive” skills scores (Fine Motor, Visual Reception, Receptive Language, and Expressive Language) which can be converted into a standardized Early Learning Composite (ELC) score. The median split-half internal consistency is above .80 for the Expressive and Receptive Language subscales, and slightly lower for the Visual Reception (.79) and Fine Motor (.75) subscales (Mullen, 1995). Test-retest reliability for the scales ranges from .71 to .96. IRR ranges from .91 to .99. Studies support the scales' convergent validity with satisfactory correlations with other measures (mean correlation of .53; Mullen, 1995). In the present study, a Mental Age (MA) score was also obtained, based on the median of the age equivalents of the four subscales (MA is the average age of the children obtaining that particular score; Sattler, 2001). The MA score was used in the main analyses in Question 2. It was also used to calculate a Ratio IQ (MA/Chronological Age x100; Sattler, 2001) in order to avoid the problem of a floor effect on low-end scores (Munson et al., 2008), as is customary in this type of research. This calculation sometimes produced very low ratio IQs (see Table 2).

Wechsler Preschool and Primary Scale of Intelligence Third Edition (WPPSI-III) (Wechsler, 2002) is a standardized, norm-referenced intelligence test designed for children ages 2 years 6 months to 7 years 3 months. The WPPSI-III provides subtest scores which represent intellectual functioning in verbal and performance cognitive domains and a composite score which represents a child's general intellectual ability (Full Scale IQ). The WPPSI-III Full Scale IQ has good reliability with internal consistency ranging from .86 to .97. The subtests have average internal consistency reliabilities that range from .83 to .95. Test-retest reliability ranges from .81 to .88 (Wechsler, 2002). Studies indicate that the WPPSI-III has satisfactory criterion validity and that it is a good measure of general intelligence (Wechsler, 2002). In the present study, Mental Age (MA) was calculated using the children's Full Scale IQ and age.

Adaptive level was measured using the Vineland Adaptive Behavior Scales, 2nd ed. (Vineland-II; Sparrow, Cicchetti, & Balla, 2005), a norm-referenced parent-interview measure of adaptive behaviour which evaluates the skills displayed in everyday situations. The results from the Vineland provide standard scores for Communication, Daily Living Skills, Socialization, and Motor domains and the Adaptive Behavior Composite (ABC). Average test-retest reliability across domains ranges from .88 to .92.

Inter-interviewer reliability for the ABC is .87, while domain reliabilities average .75 (Sparrow et al., 2005). The validity of the Vineland has been demonstrated in studies which provide evidence that the level of adaptive functioning as measured by the Vineland-II differentiates clinical groups from nonclinical groups and through its modest correlations with IQ tests such as the WISC-III (Sparrow et al., 2005).

IBI Quality was assessed using the previously-described YMQI. Briefly, it is a manualized behaviour observation measure of IBI quality in which raters code two 5-minute segments of videos of children engaged in IBI on 31 items. It is a measure with established reliability and validity based on correlations with expert judgment as described earlier. Four undergraduate coders were first trained to criterion and subsequently coded all the videos for the study. The training included studying a manual with detailed coding rules for each item, completing a quiz on behavioural principles, viewing a self-guided training DVD and completing the included exercises, and achieving at least 80% inter-rater agreement with the YMQI developers on at least three of five training videos. All coders came close to this criterion after training independently. An in-person booster training session was provided by this writer, during which one of the training videos was coded together with the coders and the criterion ratings were explained for each item. After this session, the coders re-coded the videos (not including the one used in the booster training session) and all reached acceptable reliability as outlined in the YMQI manual.

Although the goal was for videos to be made once per month for each child, for logistical reasons, this was not always the case. In total, 402 videos were coded. The number of videos per child ranged from five to 14, with a mean of 10 (median of 11) videos per child. Inter-observer agreement for 20% of these videos that were double-coded, as described earlier, ranged from 76 to 100% for individual items with a mean of 88% (Blacklock et al., 2013).

Data Analyses

Question 1

The first research question was to determine the factor structure of the YMQI empirically and to see whether it was similar to the rationally-derived subscales from a previous study (Taheri et al., 2013). In order to answer this question, an exploratory factor analysis (EFA) was performed on all the available, coded YMQI videos from the prospective study (n = 402). EFA is a statistical technique used to discover whether the pattern of correlations among a set of observed variables can be explained by a small number of latent variables, or factors. The resulting factors are thought to reflect underlying processes which create the correlations among variables.

EFA was performed on the first segment of each rated YMQI video. As previously noted, two 5-minute segments from each video are coded using the YMQI. The correlation between the total YMQI scores of the two segments was strong and significant ($r = .708$, $p < .001$). For this reason, as well as the fact that some of the videos did not have a second coded segment due to a lack of a suitable segment (e.g., not enough discriminative stimuli (S^Ds), shorter video, child playing in gym instead of engaging in IBI, etc.), only the ratings from the first coded segment were used in the EFA. Polychoric correlations were performed with this 5-point ordinal scale. A scree plot, parallel analysis, and root-mean squared residual (RMSR) statistics were used to help determine the number of factors, along with pragmatic interpretation. Enough factors were retained for an adequate fit of the final EFA model to the data; however parsimony of the solution, as well as conceptual considerations based on previous research and theory, were given equal importance. As is customary, to aid interpretation of the factor pattern, an oblique rotation was used because it allows for factors to be correlated with each other.

Question 2

Based on the results of the first question, subscale scores representing each YMQI factor were calculated and used for further analyses.

a) Because the current sample consists of the same 39 children engaged in IBI over time, whether and how the YMQI subscale scores change over the child's time in the study was explored. The level and trend of these subscales over time were examined graphically and with growth curve modeling.

b) In order to assess whether the level of the YMQI subscale scores during the child's time in IBI is a function of the child characteristics at the beginning of treatment (autism severity, cognitive level, and adaptive level), growth curve models were estimated in which the level of each subscale at the start of treatment was regressed on child characteristics at the beginning of IBI.

c) Next, the relationships of the subscale scores to children's outcomes after approximately one year of treatment (cognitive and adaptive skills, and autism severity, all at Time 2, as well as gains in cognitive and adaptive skills, and decreases in autism severity) were examined using Spearman correlations and regression models.

Results

Question 1

First, the polychoric correlations among all the items were examined (see Table 3). The polychoric correlations among the YMQI items ranged from small (e.g., .24) to large (e.g., .67).

The ordinary least squares estimation procedure was used to fit factor models to the data from the YMQI items (i.e., the polychoric correlations). The scree plot, which suggested a four-factor solution, was examined first to determine which models to explore further (see Figure 1). Parallel analysis suggested that the number of factors is nine. However, this was not a practical number for a measure with only 31 items. Previous research on the measure suggested that a possible number of dimensions was three, therefore 3-, 4- and 5-factor models were examined. Table 4 summarizes various model-data fit statistics for the three hypothesized models including Root Mean Square Residual (RMS), Root Mean Square Error of Approximation (RMSEA), Tucker Lewis Index (TLI), and Bayesian Information Criterion (BIC). Based on all model fit statistics used, the 5-factor model had the best fit to the data. Therefore, the 5-factor model was interpreted first. The rotated solution included two factors that were clearly interpretable. Two other factors were less clear conceptually and statistically, as there were several items that did not have strong factor loadings on either factor or on any of the other factors. Finally, only three items had their highest loadings on the fifth factor, two of which had very low factor loadings across all factors. The three items with their highest loadings on this fifth factor did not represent a meaningful construct, causing the 5-factor model to be uninterpretable.

Hence, the 4- and 3-factor models were examined next. Both rotated solutions were interpretable; however, the 4-factor model had better fit statistics than the 3-factor model (see Table 4; lower RMS, lower RMSEA, higher TLI, and much lower BIC). Therefore, the 4-factor model was selected and the rotated factor loadings are presented in Table 5. Oblimin rotation was used to allow the factors to be correlated with each other (r = .14 to -.22).

Upon consultation with the measure's first author, the factors were interpreted in the following way: Factor 1 is a Pace and Organization factor and refers to the flow of the session, where a more organized therapist can provide more intensive intervention; Factor 2 refers to the Technical Correctness of the IBI from a behaviour analytic perspective and includes such aspects as correct prompting and reinforcement procedures; Factor 3 is an Engagement and Motivation factor which refers to the more clinical aspect of the intervention and highlights the importance of the maintenance of the child's focus and interest in the activities being presented to them; finally Factor 4 is a Generalization factor which refers to teaching that is intended to ensure that the child is able to use his or her newly acquired skills and knowledge in different situations and environments. The correlations among the factors are presented in Table 6.

After careful consideration, Item 27 (Response Generalization) was removed from the analyses due to having very low factor loadings on all four factors and the lowest communality of all the items, indicating that only 3.09% of variation in this variable was explained by the four factors (see Table 5 for all communalities). Additionally, item 2 (Varying S^Ds) was included in the subscales for both factor 2 and factor 4. The factor loading coefficients are almost identical, although somewhat low in both cases (0.216 and

0.215 for factors 2 and 4, respectively) and the item makes conceptual sense with both factors. Item 2 fits in the generalization theme of factor 4 since varying the wording of the stimulus is a good way to avoid simple conditioned responses while working on the child's ability to answer different questions about the same concept. It also fits conceptually with factor 2, as teaching using a variety of appropriate S^Ds is considered good technical practice according to behaviour analytic principles.

After removing item 27 and including item 2 in the subscales for factors 2 and 4, subscale scores representing each of the four factors were calculated as the mean of the scores for the set of items which had their strongest loading on a given factor. The internal consistency of the four new subscales was then examined. The Pace and Organization (7 items, $\alpha = .80$), Technical Correctness (11 items, $\alpha = .74$), and Engagement and Motivation (5 items, $\alpha = .77$) subscales have good internal consistency, while the Generalization subscale (8 items, $\alpha = .61$) has acceptable internal consistency. These were the subscales used to address the research questions in the remainder of the study.

Question 2

First, the change in the four YMQI subscales during the child's participation in IBI was examined. Subscale scores from three videos were graphed: the first video represented the child's YMQI subscales at the start of IBI (the video closest to the date the child began IBI/the earliest video); the second after approximately six months of IBI (the video closest to the six-month time point); and a final video at the end of approximately one year of IBI (plus or minus 3 months, favouring videos that were filmed after at least 12 months of IBI, when possible). Each individual child's subscale scores were graphed at these three time points (see Figure 2). Each line on these graphs represents one individual child and his or her particular subscale score at each of the three time points. There appears to be much variability on all four subscales. It appears as though there is more variability in scores around the six months mark. However, the subscale levels after one year of treatment are also quite variable. This observation may be due to the fact there are fewer data points at the start of IBI, due to a lack of a video close enough to the IBI start date for several children, therefore creating less variability at Time 1. Looking at all of the graphs together, there is no discernable pattern that the subscales follow across time due to the notable variability among different children's subscale scores.

Next, the means of all the children's YMQI subscale scores at each time point were calculated and graphed for each subscale (Figure 3). This graph shows that all the subscale mean scores at all the time points are within the "good" quality range of IBI, as defined by the YMQI Administration Manual (a score from 2.1 to 2.5 is considered good quality; Perry, Flanagan, et al., 2008). All the subscale means appear to remain fairly stable across the year of treatment, with a very slight increase from the start of IBI to the six-month time point and a very slight decrease from the six-month time point to the one-year time point. Finally, similar to what was observed in the previous rationally-derived subscales, the Generalization subscale remains within the "good" quality of IBI range throughout the year, but is notably lower than the other subscales.

One major limitation of exploring the data in this way is that only three selected time points are examined, instead of all the available data points from all the YMQI videos that were recorded during approximately one year of IBI. Because many of the videos were not recorded at specific time intervals and different children have different numbers of videos, the trends in YMQI subscales over time needed to be examined in a statistically sophisticated way, specifically growth curve modeling, which allows the use of all the available data from all the time points. Growth curve modeling was selected because the procedure allows for variably spaced measurement occasions, such that each participant can have his or her own customized data collection schedule, and because it allows for varying numbers of waves of data, such that not all participants need have the same number of waves of data (Singer & Willett, 2003).

The first step in a growth curve model analysis is the identification of the optimal functional form of the trajectory over time (Curran, Obeidat, & Losardo, 2010) to determine whether the data follows a general linear or non-linear (e.g., quadratic) pattern. The data for each YMQI subscale were examined to see whether a linear or quadratic model was a better fit. The two models for each subscale were compared using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (see Bollen & Long, 1993). The AIC penalty accounts for the number of parameters in the model, while the BIC penalty also accounts for the sample size. In both cases, a smaller value indicates a better model-data fit. Likelihood ratio tests were also used to determine whether the fit of the linear and quadratic models differed significantly from one another (Table 7). For subscales 1, 2, and 4, there was no significant difference between the linear and quadratic models, supporting the fact that the AIC and BIC values were quite similar for the two models. Therefore, the simpler linear model was selected because the more complicated quadratic model did not fit the data significantly better. For subscale 3, there was a significant difference between the linear and quadratic models and the quadratic model had smaller AIC and BIC scores. Therefore, the data for three of the subscales were considered to be better fitted by the linear models, while the data for subscale 3 were better fitted by the quadratic model.

Next, separately for each subscale, the individual data for each child were graphed along with their fitted linear trajectories. All graphs are displayed in Figures 4 to 7, where each small dot represents the child's score on that particular subscale for each individual video. The fitted lines indicate whether that child's scores on the particular subscale increase, decrease, or remain fairly stable, overall, across approximately one year in IBI. As can be seen in Figures 4 to 7, there is considerable variability in the trends of subscale scores across time, such that some clearly increase, others decrease, and yet others remain relatively stable across the year. It is important to remember the lines representing the linear trend do not fully represent each particular subscale at each time point, as these are quite variable, with many points falling below and above the fitted line. Furthermore, the linear and quadratic mixed-effects model also indicates a strong negative correlation between each subscale's intercept and time (subscale 1, $r = -.85$; subscale 2, $r = -.85$; subscale 3, $r = -.73$; subscale 4, $r = -.91$). This correlation indicates that children with subscale scores that start off higher tend to decrease throughout the year, while children with scores that are low at the start of IBI tend to increase during the year.

Question 2b

Our next question was to explore the relationship between YMQI subscale scores at the start of treatment and child characteristics at the beginning of IBI, namely autism severity, cognitive level, and adaptive level. In order to minimize the number of analyses performed with the relatively small sample size, for the remainder of the analyses, one variable to represent each of these three child characteristics was selected from the various scores available from the measures used in this study. To represent cognitive level, the child's Full Scale Mental Age score was selected. This variable was selected instead of standard scores because it is less dependent on how the child performs compared to his or her same-aged peers. It is the author's belief that this variable more validly represents the child's abilities, and, importantly, it may show the most change after treatment, as it will show whether children have gained or lost skills relative to themselves instead of being compared to their typically developing peers. Since the current sample is quite low-functioning at the start of treatment (Full Scale IQ *M* at the start of IBI = 43.40), looking only at changes in standard scores may mask real gains that these children make relative to their own abilities at the start of IBI. For similar reasons, the child's Adaptive Behavior Composite Age Equivalent score to represent the children's adaptive skills was chosen. The total CARS score represents the severity of the children's autism symptomatology.

In order to answer this question, associations between initial child characteristics and subscale scores at the start of treatment were examined. In order to do so, latent growth models were estimated in which the growth factor of level of each subscale at the start of IBI was regressed on child characteristics at the start of treatment. The level of subscale 1 at the start of IBI was not predicted by any of the child variables at the start of treatment (all ps > .05; see Table 8). Similarly, the level of subscale 2 at the start of treatment was not predicted by any of children's characteristics at the beginning of IBI (all ps > .05; see Table 9). Subscale 3, however, was negatively related to CARS scores at the start of IBI, such that children with higher CARS scores (more autism symptomatology) receive lower scores on subscale 3 (Engagement and Motivation) (β = -0.01, p = 0.04; see Table 10). Finally, subscale 4 at the start of IBI was not predicted by any of the child characteristics at the beginning of treatment (all ps > .05, see Table 11).

Question 2c

Lastly, the relation of the YMQI subscale scores to children's progress in IBI, as measured by changes in cognitive and adaptive skills, as well as autism severity, was examined. First, Spearman correlations among the four subscales at the start of IBI and the amount of change children made in their cognitive and adaptive skills, as well as in autism severity scores, were examined (Table 12). To calculate the change in cognitive skills, the child's Full Scale Mental Age at Time 1 was subtracted from the child's Full Scale Mental Age at Time 2, whereby a positive number indicates gains in cognitive skills. To calculate the child's change in adaptive skills, the Adaptive Behavior Composite Age Equivalent score at Time 1 was subtracted from the Adaptive Behavior Composite Age Equivalent score at Time 2, whereby a positive number indicates gains in adaptive skills. To calculate the change in autism severity, the child's CARS total score at Time 1 was subtracted from the child's CARS total score at Time 2, where a negative number indicates an improvement in autism severity.

Subscale 1 had nonsignificant correlations with all of the children's change scores (all ps > .05). Subscale 2 was negatively correlated with the change in CARS total score (r = -.65, p = .002), indicating that IBI high in Technical Correctness (subscale 2) at the start of treatment was associated with greater decreases in autism severity. Subscale 3 had nonsignificant correlations with all of the children's change scores (all ps > .05).

Finally, subscale 4 had a negative correlation with change in CARS total (r = -.49, p = .04), and a positive correlation with change in Full Scale Mental Age scores (r = .47, p = .04), suggesting that children with IBI higher in Generalization (subscale 4) at the start of IBI had more gains in their cognitive skills and greater decreases in autism severity.

Next, the effect of the change in YMQI subscales over the child's participation in IBI, or in other words, the slope of each subscale for each child from the time they began treatment (Time 1) to approximately one year later (Time 2), on children's outcomes was examined. These effects were important to estimate because previous analyses showed that the subscales do, in fact, tend to change during the child's time in IBI. Therefore, instead of only examining the impact of IBI quality at specific time points, it was also important to explore how this change in different aspects of quality over approximately one year of IBI may impact children's outcomes. For this analysis, several sets of regression models were estimated, the first to determine whether the slope of each subscale during the child's time in IBI can predict each of: the *outcomes* at Time 2, controlling for that same variable at Time 1; and the *change* in child characteristics during the child's time in IBI. The slope of each subscale for each child was calculated using a difference score between Time 1 and Time 2 subscale levels and used in the analyses.

In the first analysis, none of the subscales' slopes predicted the outcome variables at Time 2, which were autism severity, Full Scale Mental Age, and Adaptive Behavior Composite Age Equivalent (Tables 13-16), even after controlling for the same variable at Time 1.

Following the analysis of whether the change in IBI quality over approximately one year of treatment affects children's outcomes, I also examined whether the slope of any of the subscales predicts the change in child characteristics between Time 1 and Time 2. These analyses facilitated a determination of whether any changes in skills across approximately one year of IBI can be predicted by the changes in aspects of IBI quality over the same time, as measured by the slopes of the YMQI subscales. There were no significant results for the change in children's cognitive and adaptive skills, such that none of the slopes of the four subscales predicted the change in these scores from Time 1 to Time 2 (Tables 17-20). However, the change in the CARS total was significantly predicted by the slope of subscale 2 (β = 315.19, p = .02; see Table 18), such that greater decreases in autism symptomatology were associated with greater increases in subscale 2 scores. The slope of subscale 3 also significantly predicted the change in the CARS total score (β = 200.50, p = .02; see Table 19), indicating that children with greater increases in subscale 3 during the year had greater decreases in autism symptomatology. Finally, the change in children's CARS total score was also predicted by the slope of subscale 4 (β = 200.50, p = .02; see Table 20), such that greater decreases in autism symptomatology were associated with greater increases in subscale 4.

Discussion

This study is the beginning of an examination of a very important area of IBI research which has largely been ignored in the literature, namely the quality of the treatment that children partaking in IBI receive. These data provided a unique opportunity to examine a measure of IBI quality in detail and determine its best use. Doing so allowed an initial examination of different aspects of IBI quality and whether they decrease, improve, or remain stable across one year of IBI. This study is also a first attempt to understand the relationships of different aspects of IBI quality with children's characteristics and progress in the intervention program, which have not been examined in any literature thus far. The results of this study help operationalize the quality of IBI more precisely, inform future research, have implications for IBI training and supervision, and help continue to unravel the question of why some children benefit more from IBI than others.

Question 1

In this study, one of the only measures of quality of IBI was explored in detail and the YMQI was determined to be multidimensional. As suggested by previous research and clinical experience, the YMQI indeed does not represent one unitary construct of IBI quality, but rather four related aspects of quality IBI, namely pace and organization, technical correctness, engagement and motivation, and generalization. This analysis is particularly valuable for the improvement of the measure because it helps determine which items to retain for future revisions of the YMQI and which subscales to calculate. This analysis showed that it may be best to remove item 27 (Response Generalization) from the measure in future versions, as it does not seem to measure the same constructs as the other items and is not strongly related to them statistically. This item was also problematic in previous analyses in that it was poorly correlated with total quality. In addition, although response generalization is clinically important, it may not be easily seen in a brief video and this item does not fit very well conceptually with the other items. This is the only item that is coded based on the child's responses and not solely on the behaviour of the therapist. Because the measure is meant to gauge the quality of IBI, and hence the behaviour of the therapist, and not the success of the intervention as observed in the child's responses, item 27 does not fit well with the rest of the items.

It is exciting to find that IBI quality is not solely dependent on knowledge of behavioural principles, technical proficiency, and the speed of the therapist. Engagement has been recognized as an important factor in learning for children with and without disabilities (Greenwood, 1991; McWilliam, Trivette, & Dunst, 1985). Researchers have also stated that the quality of an early intervention program, such as IBI, must be measured not only by the number of hours of the intervention, but also in terms of the effectiveness of the program in obtaining child engagement in learning opportunities, particularly for children with autism (McGee, Daly, Izeman, Mann, & Risley, 1991). Although researchers have stated that measuring and understanding engagement is a necessary step in determining how to provide high quality and effective services for students with ASD (Steinbrenner & Watson, 2015), and professionals have been highlighting the importance of clinical factors such as engagement and motivation of the child in IBI quality, this aspect has not received very much attention in IBI studies. One recent study that did examine engagement showed that children's ability to engage with therapists and other adults at the start of treatment predicted better cognitive and adaptive outcomes after one year of treatment (Smith et al., 2015). However, McGee, Morrier, and Daly (1999) state that, unfortunately, at that time most publicly funded early intervention programs lacked specialized curricula needed to promote constant engagement in children with autism, therefore reducing their impact. Results from the present study suggest that Engagement and Motivation fell in the good quality range in the current sample. The YMQI, and in particular the items that comprise the Engagement and Motivation subscale, can be used as a tool to begin to explore this important aspect of IBI in clinical settings and future research studies.

When comparing the EFA results to the previous rationally determined subscales (Taheri et al., 2013), overall, the EFA results lead to splitting up the large rationally-derived Technical Correctness subscale items into separate Technical Correctness and Pace and Organization subscales, retaining a similar Generalization subscale (containing many, though not all, of the same items), dropping the Problem Behaviour subscale, and creating an Engagement and Motivation subscale. More specifically, the rationally-derived Technical Correctness subscale included the majority of the YMQI items (21 out of 31 items). However, in the EFA results, only seven items remained in the Technical Correctness subscale. Another seven are now included in the new Pace and Organization subscale, which highlights the importance of the pace, and hence the intensity, of the treatment session. Four items also are used to calculate the new Engagement and Motivation subscale, and three are in the Generalization subscale. Five of the items from the rationally-derived Generalization subscale were retained in the EFA-derived Generalization subscale. Finally, the EFA did not suggest a separate Problem Behaviour subscale, and the items from this rationally-derived subscale are now used to calculate the Technical Correctness and Engagement and Motivation subscales. Notably, the EFA highlighted the importance of the maintenance of the child's focus and interest in the activities in the Engagement and Motivation subscale, which was not part of the rationally-derived subscales. The conclusions reached from the EFA incorporate both statistical and clinical interpretation of the results from the author of this study as well as the YMQI's first author, thus making them sounder than simply the rationally-derived subscales. Future research is needed to cross validate the factor structure of the YMQI further by performing a confirmatory factor analysis on an independent sample of YMQI ratings to test the theory about latent processes of IBI quality and confirm the current results.

The four-factor model of the YMQI has important implications for evaluating the quality of intervention that thousands of children receive in private and publicly-funded clinical settings. This finding can be used to assess performance, provide guidance, evaluate existing programs, and train current and future IBI therapists to ensure the program's strong performance with all aspects of IBI quality. This finding is also important for future research and validates a means whereby the quality of IBI can be explored in forthcoming studies. For example, one recent study used several measures of competencies, including the YMQI, of tutors working in a school for children with autism in the UK (Denne, Thomas, Hastings, & Hughes, 2015). Their results supported the criterion validity of the measures, in that the more experienced tutors achieved higher scores on measures they used, including the YMQI, although the measures were not strongly correlated with each other. Another ongoing study is using the YMQI to give tailored supervision to service providers at the Autism Center in Stockholm and to estimate the correlation between high quality teaching, as measured by the YMQI, and the participating children's learning rate (Långh, 2014). Yet another recent study (Foran & Hoerger, 2014) used the YMQI to measure quality of intervention in a low intensity behavioural treatment ABA program in a special needs school for children with ASD and learning disabilities in the UK, and showed that the program can be effective in achieving gains in cognitive and adaptive skills despite the limited number of hours during which it was delivered, presumably due to the high quality teaching as measured by the YMQI. The use of the YMQI in these studies is particularly notable since the YMQI is being practically applied to behavioural programs outside of North America. Despite cultural differences, such as differing social reinforcement styles, and traditional educational practices that may include the provision of negative attention and low amounts of praise (as those outlined in Jones et al., 2011), the YMQI is finding applicability in a variety of contexts and cultures. Research questions such as these, as well as many others, will be able to be addressed with more targeted precision in terms of the different aspects of behavioural intervention quality and can therefore have a stronger impact on personnel training and evaluation and on the examination of the relationship between quality behavioural teaching and children's outcomes.

Question 2

An important finding is that participating children from the sites in this study received good quality IBI overall, as measured and defined by the YMQI. It is important to note that, although these videos represent the IBI quality that children actually received in the real world, the number of therapists and videos varied across children. This limitation is important to keep in mind when interpreting the results of this study as change in quality over time. Ideally, the data would have the same number of videos and the same therapists working with each participating child consistently throughout the year, however this was not possible in a real-world IBI program. More specifically, the examination of how the quality of IBI changes over approximately one year of treatment showed that there was substantial variability on all subscales among the children. However, when looking at the means of the subscales across time, it was encouraging that all aspects of IBI quality were in the "good" quality range throughout the year. Notably, the Generalization subscale was the lowest, although still within the "good" range, across all three time points. This Generalization result could occur as an artifact of the way this subscale is measured by the YMQI items, most of which are "evidence"-type items, but it could also represent what actually happens in real-world IBI programs. For example, these programs might focus more on the other aspects of IBI quality during the first year compared to the focus on generalization, which may be focused on later in treatment. In addition, it may be difficult to create opportunities for generalization before the child has mastered other important objectives, as it is typical to acquire skills first and then work on generalizing them. Finally, as generalization often occurs away from the table (e.g., in the gym or staff offices), it may have been difficult to observe these opportunities in the 5-minute video segments. The level of Generalization that individual children in the current study actually received was quite variable and it will be important to revisit this finding in future research, as well as in clinical settings, as this study highlights the importance of generalization across settings and persons in order to improve children's skills, as advocated in the ABA literature for many years (Stokes & Baer, 1977).

To examine the question of quality over time further, a more sophisticated statistical analysis, namely growth curve modeling, was used, which was forgiving to the problems in the data, such as different numbers of subscale scores available for different children and uneven periods of time between the videos. The overall pattern of the subscale scores across time was determined to be linear for three of the subscales, and quadratic for subscale 3. Furthermore, for each subscale, scores that were high at the beginning of IBI tended to decrease during the child's time in IBI, while scores that were low at the start of treatment tended to increase over the year. This finding could simply reflect regression to the mean, a tendency for extreme scores to become less so on repeated measurement. More meaningfully, this strong relationship could reflect that therapists who start out with lower quality scores may improve in their abilities with practice and supervision throughout the year, while therapists who are very skilled at the start of IBI may become more lenient in their delivery of the intervention over time. This decline could happen for many reasons, such as unmeasured child and therapist characteristics which were not accounted for in this study.

Ultimately, IBI quality, especially the way in which it was measured in the current study, is reliant on the IBI therapists' delivery of the intervention (Denne, Hastings, Hughes, Bovelic, & Redford, 2011). As Griffith, Barbakou, and Hastings (2014) state, "delivering ABA therapy is an intensive, repetitive task, often on a one-to-one basis, with children who may engage in intense and/or frequent challenging behaviours" (p. 549), which can be quite taxing on the therapists. Symes et al. (2006) asked 19 therapists providing behavioural instruction to young children with autism in the home to identify factors that facilitated or impeded their capacity to deliver the intervention. The therapists identified several factors that were not measured in this study, such as therapist factors (their own patience), the quality of their training, and child factors. In particular, child factors such as compliance and competence were considered to facilitate instruction by the therapists, while child challenging behaviours and lack of progress were reported to hinder therapists' procedural fidelity (Symes et al., 2006). It is therefore likely that many factors that were not taken into account in this study may have caused these modest decreases in IBI quality (although still remaining in the "good" quality range overall) in some children during the first year of intervention. However, this possibility is somewhat mitigated by the fact that different therapists worked with the children across videos.

Our examination of whether children with different characteristics when they begin IBI might receive treatment focused on different aspects of quality showed that subscales representing Pace and Organization, Technical Correctness, and Generalization were not predicted by any of the initial child characteristics. However, the Engagement and Motivation subscale was predicted by the child's CARS score, such that children with more severe autism symptomatology engaged in IBI that was rated lower on the Engagement and Motivation subscale, while children with less autism symptomatology engaged in IBI higher in Engagement and Motivation at the start of the intervention. This result is not surprising, as a lack of engagement, in particular social engagement, has been proposed to be a core deficit underlying the development of autism (Hobson, 1993; Rogers & Pennington, 1991). Similarly, the lack of social motivation has been proposed as a core deficit of autism (Berger, 2006; Grelotti, Gauthier, & Schultz, 2002; Klin, Jones, Schultz, & Volkmar, 2003; Koegel & Koegel, 1995). Early deficits in social motivation in children with autism have been suggested to lead to a lack of orientation and engagement in the environment (Volkmar & Klin, 2005), perhaps underlined by neurological dysfunction (Berger, 2006). It is not surprising that children who are more "severely autistic" at the start of IBI have subscale scores which are lower in Engagement and Motivation at the beginning of treatment because these children are likely much more difficult to engage in the therapy and may have more difficulty with the social aspect of the one-to-one intervention and reinforcement. This relationship validates that subscale 3 of the YMQI does indeed measure Engagement and Motivation in therapy.

Our last analyses examined the relationships of the YMQI subscales to children's outcomes in IBI, which were defined as the change in children's cognitive and adaptive skills, and autism severity over approximately one year of IBI. Pace and Organization of the IBI at the start of treatment was not related to any gains in skills or decreases in autism severity, which is somewhat surprising as the intensity of the intervention is viewed as a hallmark of IBI. However, in the literature, intensity is often measured by the number hours of intervention that the child receives and not necessarily the number of trials during an IBI session, which is what the YMQI is measuring. The other YMQI subscales had some notable relationships with changes in the children's abilities. In particular, children who had greater decreases in their autism symptomatology had received good quality IBI that was higher in Technical Correctness and Generalization at the start of treatment. This result shows that IBI that is higher quality overall is related to bigger decreases in children's autism symptomatology. This may be because, in the first year of treatment, IBI programs target decreasing the child's autism symptoms so that children become more engaged in their learning.

The children who made the most gains in their cognitive skills had IBI higher in Generalization at the start of treatment, which may be because of a particular characteristic of the children (e.g., stronger verbal abilities) that made them benefit from more flexible and generalizable teaching instead of IBI focused more on Technical Correctness. The change in adaptive behaviour was not related to the quality of IBI at the start of treatment, possibly because of the characteristics of the children or because of the skills targeted during the first year of treatment. This would not be surprising as adaptive skills are not the primary goal of IBI. Further, in the published literature, adaptive skills tend to change less than IQ scores.

Finally, I also examined whether changes in IBI quality during the year, as measured by the slopes of the YMQI subscales, could predict children's outcomes after approximately one year of IBI, as measured by cognitive and adaptive skills, and autism severity. There were no significant relations between the subscale slopes and children's outcomes, even while controlling for those same child variables at Time 1 for each outcome. As outlined in the Introduction, children's outcomes after one year of IBI are quite variable and researchers are still attempting to determine which factors lead to this variability, including child, family, and intervention variables. The current study's analyses included some, though not all possible, child factors, no family factors, and only the quality of the direct intervention as the intervention factor (and not other aspects of quality such as supervision, curriculum programming, and organizational and administration factors, nor any therapist factors). A combination of these factors would lead to better prediction of children's outcomes and should be explored systematically in future studies.

In addition, the change in children's cognitive and adaptive skills was not predicted by the slopes of the YMQI subscales; however, the decrease in children's autism severity was. In particular, IBI quality that increases in Technical Correctness, Engagement and Motivation, and Generalization, during the child's year of participation in IBI, significantly predicted children's decreases in autism symptomatology. This finding shows that IBI quality, overall, leads to bigger decreases in autism symptomatology. It is important to keep in mind that these data are only for approximately the first year of IBI, and therefore, IBI quality could predict gains in other skills during later intervention. It is possible then, that the first year of IBI focuses specifically on the deficits and behavioural problems of children with ASD, thus preparing them to learn at a more typical pace in future interventions and settings. In fact, clinical experience suggests that children must first be taught to sit, attend to a therapist, and actively engage in the therapy.

This study has several limitations. One major limitation is the relatively small sample size for the complex analyses that were performed. This issue is somewhat alleviated by the large number of videos that were used for several of the analyses; however, it is still a notable limitation which could lead to some Type II errors in the results. In particular, it is important to note that all of the data are nested within the same 39 children taking part in IBI in Ontario over a one-year period. This limitation is particularly important when interpreting the factor structure of the measure, as it may cause some items to be more similarly rated, and therefore affect the EFA results, than if all the videos belonged to different children.

Another limitation is the length of the study. More specifically, the data from this study cover approximately 15 months of IBI, and all of the results are based on what occurs during this time. This aspect is particularly important to keep in mind when looking at children’s outcomes, as they may change as the children continue to participate in the intervention. Perry et al. (2016) reviewed studies examining this question. For example, Sallows and Graupner (2005) reported that, although the greatest gains occurred during the first year of treatment, some children improved with ongoing treatment up to 4 years in length. In addition, Ben-Itzchak et al. (2014) reported an increase in children's social skills during the second year of treatment. These studies highlight the fact that children in the current sample may continue to make gains past the Time 2 assessment and, therefore, it is important to remember that the outcomes presented in this study do not comprehensively represent the children's "outcomes" after completing IBI, but rather represent the children's scores at an assessment approximately one year after they began treatment. In fact, ongoing learning, although at less accelerated rates, could be anticipated in a structured behavioural intervention.

It is also important to note that the data from this study come from three particular centres providing government-funded IBI in Ontario, with most of the data coming from one of those centres. Although some variability in IBI quality is observed in the current study, the treatment quality may not be representative of the quality of IBI being delivered in the wider world (including in private settings). The participating children were consecutive referrals and were not biased in any systematic way. Families were not randomly selected from a large pool of participants; rather, they were asked to participate in this study at their assessment prior to entry into IBI, with a very low rate of families declining. The sampling may limit the representativeness of these results to other publicly-funded and private IBI programs within Ontario and elsewhere. A major limitation of this study is the lack of a comparison group of similar children who receive a different treatment or no treatment. Therefore, the changes observed in children's skills during this study cannot conclusively be attributed to the IBI or its quality.

Some limitations are also inherent in the clinical nature of this study and the measures used. For example, two different IQ tests were used based on their clinical appropriateness, deviation IQs and ratio IQs were combined, and change in age equivalent scores were used which have significant psychometric limitations. However, these limitations are common to most other studies using similar measures. As previously mentioned, several other variables which can affect children's progress in IBI were not accounted for in this study due to its focus on intervention quality, such as child (e.g., age, problem behaviour, specific language abilities) and family characteristics (e.g., socio-economic status, parental engagement in the therapy). For example, the children in the current sample were older on average (M = 4 years, 10 months) than the recommended age for starting IBI. Providing IBI outside the ideal 'sensitive period' may have had an effect on the results. Finally, the researchers, clinicians, and students that conducted the assessments and recorded the videos were not blind to the children's participation in IBI, and in some cases, were not independent of the organizations providing the IBI. However, the YMQI coders who rated the quality of the IBI videos were independent of these organizations.

Despite these limitations, this is an essential first examination of the different aspects of IBI quality, which has been largely unexplored in the literature. It is important to address this gap to assess and monitor the quality of this treatment and to help explain children's variable outcomes after IBI. Although others have proposed that quality is important, this study provides significant evidence that specific aspects of the quality of the intervention are important for children's progress in IBI. This research can greatly affect the delivery of IBI in public and private domains, and can be used to help to train therapists and ensure that children receive high quality treatment that leads to better outcomes and has a positive impact on the children’s and their families’ lives.

References

Allen, K. D., & Warzak, W. J. (2000). The problem of parental nonadherence in clinical behavior analysis: Effective treatment is not enough. *Journal of Applied Behavior Analysis*, *33*(3), 373-391. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619543972?accountid=15182

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based early intervention with autistic children. *Education & Treatment of Children*, *10*(4), 352-366. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/617577595?accountid=15182

Beglinger, L., & Smith, T. (2005). Concurrent validity of social subtype and IQ after early intensive behavioral intervention in children with autism: A preliminary investigation. *Journal of Autism and Developmental Disorders, 35*(3), 295-303. doi:http://dx.doi.org/10.1007/s10803-005-3292-3

Ben-Itzchak, E., Watson, L. R., & Zachor, D. A. (2014). Cognitive ability is associated with different outcome trajectories in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(9), 2221-2229. doi:http://dx.doi.org/10.1007/s10803-014-2091-0

Berger, M. (2006). A model of preverbal social development and its application to social dysfunctions in autism. *Journal of Child Psychology and Psychiatry, 47*(3-4), 338-371. doi:http://dx.doi.org/10.1111/j.1469-7610.2005.01586.x

Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2002). "Progress and outcomes for children with autism receiving parent-managed intensive interventions": Erratum. *Research in Developmental Disabilities, 23*(1), 79-104. doi:http://dx.doi.org/10.1016/S0891-4222(02)00094-X

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch early intervention program after 2 years. *Behaviour Change, 10*(2), 63-74. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/618469082?accountid=15182

Blacklock, K., Perry, A., & Whiteford, S. (2011, May). *Current research on the York Measure of Quality of Intensive Behavioural Intervention*. Presentation at the Association for Behavior Analysis International, Denver, CO.

Blacklock, K., Shine, R., & Perry, A. (2013, April). *Reliability of the York Measure of Quality of Intensive Behavioural Intervention*. Poster presented at the 21st Ontario Association on Developmental Disabilities Research Special Interest Group Conference, Niagara Falls, ON.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics, 27,* S145-S155.

doi:http://dx.doi.org/10.1097/00004703-200604002-00013

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/742875834?accountid=15182

Davis, B. J., Smith, T., & Donahoe, P. (2002). Evaluating supervisors in the UCLA treatment model for children with autism: Validation of an assessment procedure. *Behavior Therapy, 33*(4), 601-614. doi:http://dx.doi.org/10.1016/S0005-7894(02)80019-3

Denne, L. D., Hastings, R. P., Hughes, J. C., Bovelic, V., & Redford, L. (2011). Developing a competence framework for ABA and autism: What can we learn from others? *European Journal of Behavior Analysis, 12*(1), 217-230. Retrieved from http://www.researchgate.net/publication/230634004_Developing_a_competence_framework_for_ABA_and_Autism_What_can_we_learn_from_others

Denne, L. D., Thomas, E., Hastings, R. P., & Hughes, J. C. (2015). Assessing competencies in applied behavior analysis for tutors working with children with autism in a school-based setting. *Research in Autism Spectrum Disorders, 20,* 67-77. doi:http://dx.doi.org/10.1016/j.rasd.2015.08.008

Dunn Geier, J., Perry, A., & Freeman, N. L. (2012, November). *Multi-site controlled study of the effectiveness of IBI for children with autism in community settings.* Presentation at the Ontario Association for Behaviour Analysis Conference, Toronto, ON.

Eikeseth, S., Klintwall, L., Jahr, E., & Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders, 6*(2), 829-835. doi:http://dx.doi.org/10.1016/j.rasd.2011.09.002

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism: A 1-year comparison controlled study. *Behavior Modification, 26*(1), 49-68. doi:http://dx.doi.org/10.1177/0145445502026001004

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification, 31*(3), 264-278. doi:http://dx.doi.org/10.1177/0145445506291396

Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders, 36*(2), 211-224. doi:http://dx.doi.org/10.1007/s10803-005-0058-x

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

doi:http://dx.doi.org/10.1080/15374410902851739

Eldevik, S., Hastings, R. P., Jahr, E., & Hughes, J. C. (2012). Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *Journal of Autism and Developmental Disorders, 42*(2), 210-220. doi:http://dx.doi.org/10.1007/s10803-011-1234-9

Elsabbagh, M., Divan, G., Koh, Y., Kim, Y. S., Kauchali, S., Marcín, C., . . . Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research, 5*(3), 160-179. doi:http://dx.doi.org/10.1002/aur.239

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis & Intervention in Developmental Disabilities, 5*(1-2), 49-58. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/617116993?accountid=15182

Flanagan, H. E., Perry, A., & Freeman, N. L. (2012). Effectiveness of large-scale community-based intensive behavioral intervention: A waitlist comparison study exploring outcomes and predictors. *Research in Autism Spectrum Disorders, 6*(2), 673-682. doi:http://dx.doi.org/10.1016/j.rasd.2011.09.011

Foran, D., & Hoerger, M. L. (2014). An Evaluation of a Low-Intensity, High-Quality Behavioural Intervention for Children With ASD. Paper presented at the *Association for Behavior Analysis International 40th Annual Convention*, Chicago, IL.

Freeman, N., & Perry, A. (2010). Outcomes of intensive behavioural intervention in the Toronto preschool autism service. *Journal on Developmental Disabilities, 16*(2), 17-32. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/888750958?accountid=15182

Goin-Kochel, R., Myers, B., Hendricks, D., Carr, S., & Wiley, S. (2007). Early responsiveness to behavioral intervention predicts outcomes among preschool children with autism. *International Journal of Disability, Development and Education, 54,* 151-175. doi:http://dx.doi.org/10.1080/10349120701330404

Granpeesheh, D., Kenzer, A., & Tarbox, J. (2011). Comparison of two-year outcomes for children with autism receiving high versus low-intensity behavioral intervention. *European Psychiatry, 26,* 1839. doi:http://dx.doi.org/10.1016/S0924-9338(11)73543-X

Green, G. (1996). Early behavioral intervention for autism: What does research tell us? *Behavioral intervention for young children with autism: A manual for parents and professionals.* (pp. 29-44) PRO-ED, Austin, TX. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619033443?accountid=15182

Greenwood, C. R. (1991). Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children, 57*(6), 521-535. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/617956383?accountid=15182

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology, 40*(3), 213-225. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/619868234?accountid=15182

Griffith, G. M., Barbakou, A., & Hastings, R. P. (2014). Coping as a predictor of burnout and general health in therapists working in ABA schools. *European Journal of Special Needs Education, 29*(4), 548-558. doi:http://dx.doi.org/10.1080/08856257.2014.952915

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *The Quantitative Methods for Psychology, 8*(1), 23-34. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four- to six-year follow-up. *Journal of Autism and Developmental Disorders, 30*(2), 137-142. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619487241?accountid=15182

Hartmann, D. P., & Wood, D. D. (1990). Observational methods. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy (2nd ed.)* (107-138). New York, NY: Plenum Press.

Hayward, D., Eikeseth, S., Gale, C., & Morgan, S. (2009). Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism, 13*(6), 613-633.

doi:http://dx.doi.org/10.1177/1362361309340029

Hobson, R. P. (1993). The emotional origins of social understanding. *Philosophical Psychology, 6*(3), 227-249. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/618450986?accountid=15182

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*(4), 359-383. doi:http://dx.doi.org/10.1016/j.ridd.2004.09.005

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*(1), 23-41. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/621978710?accountid=15182

Hundert, J., Walton-Allen, N., Earle-Williams, K., Sim, M., & Cope-Scott, K. (2000). *Intensive Behavioural intervention: A manual for instructor therapists.* Hamilton, ON: Behaviour Institute.

Jacobson, J. W. (2000). Early intensive behavioral intervention: Emergence of a consumer-driven service model. *The Behavior Analyst, 23*(2), 149-171. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619540115?accountid=15182

Johnston, J. M., & Sherman, R. A. (1993). Applying the least restrictive alternative principle to treatment decisions: A legal and behavioral analysis. *The Behavior Analyst, 16*(1), 103-115. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/618342816?accountid=15182

Jones, E. W., Hoerger, M., Hughes, J. C., Williams, B. M., Jones, B., Moseley, Y., . . . Prys, D. (2011). ABA and diverse cultural and linguistic environments: A Welsh perspective. *Journal of Behavioral Education*, *20*(4), 297-305. doi:http://dx.doi.org/10.1007/s10864-011-9138-5

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. In U. Frith, & E. Hill (Eds.), *Autism: Mind and brain.* (pp. 127-159) Oxford University Press, New York, NY. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/620355333?accountid=15182

Klintwall, L., Eldevik, S., & Eikeseth, S. (2015). Narrowing the gap: Effects of intervention on developmental trajectories in autism. *Autism*, *19*(1), 53-63. doi:http://dx.doi.org/10.1177/1362361313510067

Klintwall, L., Gillberg, C., Bölte, S., & Fernell, E. (2012). The efficacy of intensive behavioral intervention for children with autism: A matter of allegiance? *Journal of Autism and Developmental Disorders, 42*(1), 139-140. doi:http://dx.doi.org/10.1007/s10803-011-1223-z

Koegel, L. K., & Koegel, R. L. (1995). Motivating communication in children with autism. *Learning and cognition in autism.* (pp. 73-87) Plenum Press, New York, NY. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/618759301?accountid=15182

Koegel, R. L., Russo, D. C., & Rincover, A. (1977). Assessing and training teachers in the generalized use of behavior modification with autistic children. *Journal of Applied Behavior Analysis, 10*(2), 197-205. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/616217014?accountid=15182

Långh, U. (2014, May). The Level of Knowledge and Allegiance Among Preschool Trainers. Paper presented at the *Association for Behavior Analysis International 40th Annual Convention*, Chicago, IL.

Leaf, R. B., & McEachin, J. J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism.* New York, NY: DRL Books.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1), 3-9. doi:http://dx.doi.org/10.1037/0022-006X.55.1.3

Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques.* Austin, TX: Pro-Ed.

MacDonald, R., Parry-Cruwys, D., Dupere, S., & Ahearn, W. (2014). Assessing progress and outcome of early intensive behavioral intervention for toddlers with autism.

Research in Developmental Disabilities, 35(12), 3632-3644. doi:http://dx.doi.org/10.1016/j.ridd.2014.08.036

Magiati, I., Charman, T., & Howlin, P. (2007). A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 48*(8), 803-812. doi:http://dx.doi.org/10.1111/j.1469-7610.2007.01756.x

Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autistic spectrum disorders. *Research in Autism Spectrum Disorders, 4*(4), 577-593. doi:http://dx.doi.org/10.1016/j.rasd.2010.01.014

Malcolm-Smith, S., Hoogenhout, M., Ing, N., Thomas, K. G. F., & de Vries, P. (2013). Autism spectrum disorders: Global challenges and local opportunities. *Journal of Child and Adolescent Mental Health, 25*(1), 1-5. doi:http://dx.doi.org/10.2989/17280583.2013.767804

McGee, G. G., Daly, T., Izeman, S. G., Mann, L. H., & Risley, T. R. (1991). Use of classroom materials to promote pre-school engagement. *Teaching Exceptional Children, 23*(4), 43-47. Retrieved from http://www.researchgate.net/publication/234748189_Use_of_Classroom_Materials_to_Promote_Preschool_Engagement

McGee, G. G., Morrier, M. J., & Daly, T. (1999). An incidental teaching approach to early intervention for toddlers with autism. *Journal of the Association for Persons with Severe Handicaps, 24*(3), 133-146. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619431098?accountid=15182

McWilliam, R. A., Trivette, C. M., & Dunst, C. J. (1985). Behavior engagement as a measure of the efficacy of early intervention. *Analysis & Intervention in Developmental Disabilities, 5*(1-2), 59-71. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/617117935?accountid=15182

Ministry of Community and Social Services (MCSS). (2000). *Program guidelines for the Ontario Intensive Early Intervention Program.* Toronto, ON: Author.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., . . . Abbott, R. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. *American Journal on Mental Retardation, 113*(6), 439-452. doi:http://dx.doi.org/10.1352/2008.113:439-452

Office of the Auditor General of Ontario. (2013). *2013 Annual Report.* Ottawa, ON: Queen's Printer for Ontario. Retrieved from http://www.auditor.on.ca/en/reports_en/en13/2013ar_en_web.pdf

Penn, H. (2005). *The validity of the York Measure of Quality of Intensive Behavioural Intervention.* Unpublished Master's thesis, York University, Toronto, ON.

Penn, H. E., Prichard, E. A., & Perry, A. (2007). The reliability and validity of a pilot version of the York Measure of Quality of Intensive Behavioural Intervention. *Journal on Developmental Disabilities, 13*(3), 149-166. Retrieved from http://www.oadd.org/publications/journal/issues/vol13no3/download/pennEtAl.pdf

Perry, A. (2002). Intensive early intervention program for children with autism: Background and design of the Ontario preschool autism initiative. *Journal on Developmental Disabilities, 9*(2), 121-128. Retrieved from http://www.oadd.org/index.php?page=414

Perry, A., Blacklock, K., & Dunn Geier, J. (2013). The relative importance of age and IQ as predictors of outcomes in intensive behavioral intervention. *Research in Autism Spectrum Disorders, 7*(9), 1142-1150. doi:http://dx.doi.org/10.1016/j.rasd.2013.06.004

Perry, A., Condillac, R. A., Freeman, N. L., Dunn Geier, J., & Belair, J. (2005). Multi-site study of the childhood autism rating scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders, 35*(5), 625-634. doi:http://dx.doi.org/10.1007/s10803-005-0006-9

Perry, A., Cummings, A., Dunn Geier, J. D., Freeman, N. L., Hughes, S., LaRose, L., . . . Williams, J. (2008). Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 2*(4), 621-642. doi:http://dx.doi.org/10.1016/j.rasd.2008.01.002

Perry, A., Cummings, A., Dunn Geier, J. D., Freeman, N. L., Hughes, S., Managhan, T., . . . Williams, J. (2011). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 5*(1), 592-603. doi:http://dx.doi.org/10.1016/j.rasd.2010.07.003

Perry, A., Flanagan, H.E., & Prichard, E. A. (2008). *Administration Manual.* Unpublished Manual. York University, Toronto, ON.

Perry, A., & Freeman, N.L. (1996, November). *Outcomes of early behavioural intervention for autism: What does the research tell us?* Presentation at the Ontario Association for Applied Behaviour Analysis Conference, Toronto, ON.

Perry, A., Koudys, J., & Blacklock, K. (2016). Early intensive behavioural intervention and training. In N.N. Singh (Ed.), *Clinical Handbook of Evidence-based Practices for Individuals with Intellectual Disabilities.* Springer. DOI: 10.1007/978-3-319-26583-4.

Perry, A., Prichard, E. A., & Penn, H. E. (2006). Indicators of quality teaching in intensive behavioral intervention: A survey of parents and professionals. *Behavioral Interventions, 21*(2), 85-96. doi:http://dx.doi.org/10.1002/bin.212

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1), 60-69. doi:http://dx.doi.org/10.1016/j.rasd.2010.03.011

Prichard, E. A. (2005). *The reliability of the York Measure of Quality of Intensive Behavioural Intervention.* Unpublished Master's thesis, York University, Toronto, ON.

Prichard, E. A., Penn, H. E., Perry, A., Solish, A., & Levy, A. (2007, March). *The reliability and validity of the York Measure of Quality of Intensive Behavioural Intervention.* Poster presentation at the Society for Research in Child Development, Boston, MA.

Provincial Regional Trainers Network, Ministry of Community, Family and Children's Services (MCFCS), (2004). *Provincial IBI competencies evaluation form: Draft.*

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*(4), 512-520. doi:http://dx.doi.org/10.1007/s10803-011-1218-9

Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for your children with autism based on the UCLA young autism project model. *Journal of Autism Developmental Disorders, 39*, 23-41. Retrieved from http://link.springer.com/article/10.1007%2Fs10803-008-0596-0

Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., . . . Ward, N. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation, 112*(6), 418-438. doi:http://dx.doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Rogers, S. J., & Pennington, B. F. (1991). A theoretical approach to the deficits in infantile autism. *Development and Psychopathology, 3*(2), 137-162. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/618076688?accountid=15182

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

Retardation, 110(6), 417-438. doi:http://dx.doi.org/10.1352/0895-8017(2005)110[417:IBTFCW]2.0.CO;2

Sattler, J. M. (2001). *Assessment of children: Cognitive applications (4th ed.).* Jerome M Sattler Publisher, La Mesa, CA. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619673929?accountid=15182

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). The Childhood Autism Rating Scale (CARS). Los Angeles, CA: Western Psychological Services.

Sheinkopf, S. J., & Siegel, B. (1998). Home based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders, 28*(1), 15-23. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619310178?accountid=15182

Shook, G. L. (1993). The professional credential in behavior analysis. *The Behavior Analyst, 16*(1), 87-101. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/618345334?accountid=15182

Shook, G. L., Hartsfield, F., & Hemingway, M. (1995). Essential content for training behavior analysis practitioners. *The Behavior Analyst, 18*(1), 83-91. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/618691764?accountid=15182

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press, New York, NY. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619994650?accountid=15182

Smith, T., Groen, A. D., & Wynn, J. W. (2000). "Randomized trial of intensive early intervention for children with pervasive developmental disorder": Erratum. *American Journal on Mental Retardation, 105*(6), 508. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/619735191?accountid=15182

Smith, T., Klorman, R., & Mruzek, D. W. (2015). Predicting outcome of community-based Early Intensive Behavioral Intervention for children with autism. *Journal of Abnormal Child Psychology, 43,* 1271-1281. doi:http://dx.doi.org/10.1007/s10802-015-0002-2

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Circle Pines, MN: AGS Publishing.

Steinbrenner, J. R. D., & Watson, L. R. (2015). Student engagement in the classroom: The impact of classroom, teacher, and student factors. *Journal of Autism and Developmental Disorders, 45*(8), 2392-2410. doi:http://dx.doi.org/10.1007/s10803-015-2406-9

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349-367. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/616217809?accountid=15182

Symes, M. D., Remington, B., Brown, T., & Hastings, R. P. (2006). Early intensive behavioral intervention for children with autism: Therapists' perspectives on achieving procedural fidelity. *Research in Developmental Disabilities, 27*(1), 30-42. doi:http://dx.doi.org/10.1016/j.ridd.2004.07.007

Tabachnick, B., & Fidell, L. (2007). *Using Multivariate Statistics (5th ed.)*. Boston, MA: Pearson Education Inc.

Taheri, A., Blacklock, K., & Perry, A. (2013, November). *Is IBI Treatment Quality Consistent Over Time?* Poster presented at the 2013 Ontario Association for Behaviour Analysis, Toronto, ON.

Virues-Ortega, J., Rodríguez, V., & Yu, C. T. (2013). Prediction of treatment outcomes and longitudinal analysis in children with autism undergoing intensive behavioral intervention. *International Journal of Clinical and Health Psychology, 13*(2), 91-100. doi:http://dx.doi.org/10.1016/S1697-2600(13)70012-7

Volkmar, F. R., & Klin, A. (2005). Issues in the classification of autism and related conditions. *Handbook of autism and pervasive developmental disorders, vol. 1: Diagnosis, development, neurobiology, and behavior (3rd ed.).* (pp. 5-41). Hoboken, NJ: Wiley. Retrieved from http://search.proquest.com.ezproxy.library.yorku.ca/docview/620806537?accountid=15182

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence-III*. San Antonio, TX: Psychological Corporation.

Whiteford, S., Blacklock, K., & Perry, A. (2012). Brief report: Reliability of the York Measure of Quality of IBI (YMQI). *Journal on Developmental Disabilities*, *18*(3), 63-67. Retrieved from http://ezproxy.library.yorku.ca/login?url=http://search.proquest.com/docview/1440032603?accountid=15182

Wood, W. S. (Ed.). (1975). *Issues in evaluating behavior modification.* Champaign, IL: Research Press.

Table 1

Categories, Descriptions, and Individual Items of the YMQI

Category	Description	Items
A. Discriminative Stimuli	Verbal and nonverbal instructions given to the child to follow	1. Attending during S^Ds 2. Varying S^Ds
B. Reinforcement	Stimulus that strengthens a response it follows, such that the response is more likely to occur again in the future, and includes verbal praise or feedback, tangible items, social/physical activities, games, or tokens, as well as negative reinforcement such as a break	3. Rapid reinforcer delivery 4. Motivating reinforcers 5. Varying reinforcers 6. Relation of reinforcers to the task 7. Sincere/motivating verbal reinforcers 8. Differential reinforcement
C. Prompting	Stimulus that increases the probability of a correct response and can follow or precede the S^D, and may be full or partial, physical, modeled, gestural, or verbal	9. Effectiveness of prompts 10. Fading and augmenting of prompts 11. Lack of prompting errors 12. Follow through 13. Implementation of error correction
D. Organization	Preparation of the therapist in terms of knowing what she is supposed to be teaching and having all necessary materials accessible and arranged properly to facilitate the session running smoothly	14. Clear plan and teaching goals 15. Accessible materials
E. Pacing	Speed at which trials are presented and the rate at which the child is exposed to learning opportunities	16. Length of inter-trial intervals 17. Suitable pace for the child 18. Intensive teaching
F. Teaching Level	Level of task difficulty that is appropriate for the child, such that there is evidence of learning new skills and making progress	19. Suitable task difficulty 20. Evidence of skill acquisition
G. Instructional Control	Therapist's ability to engage the child for a significant period of time without the child exhibiting any problem behaviours	21. On-task following requests 22. Maintenance of the child's focus

H. Generalization	Ability to generalize the skills learned in therapy to new situations, including stimulus generalization, response generalization, child-directed learning, teaching away from the table, and capitalizing on teachable moments	23. Varying teaching materials 24. Mixing tasks 25. Teaching away from the table 26. Teaching embedded in naturalistic activities 27. Response generalization 28. Flexible teaching
I. Problem Behaviour	Includes serious behaviour (e.g., self-injury, aggression, destruction, and having a tantrum), and less serious behaviours (e.g., throwing teaching materials, refusing to participate or return to task, repetitive behaviour), and less severe behaviours (e.g., crying, being silly, getting up from the table when not allowed); problem behaviour is best thought of in terms of the function it serves for the child, as the same behaviour may serve different functions (e.g., access to tangibles, attention, sensory reinforcement, escape/avoidance)	29. Result of problem behaviour 30. Reinforcement of appropriate behaviour 31. Use of prevention strategies

Table 2

Developmental and Diagnostic Status at Start of IBI (Time 1) for Sample Used in Analyses (n = 39)

	Range	*M* (*SD*)
Gender	Male; 36 (92.3%)	
	Female; 3 (7.7%)	
Age	40 – 87 months	58.51 (12.43)
Diagnosis	Autistic Disorder; 30 (76.9%)	
	PDD-NOS*; 9 (23.1%)	
CARS (Total)	24 – 48	33.79 (6.10)
Cognitive level		
FSIQ (*n* = 38)	16.67 – 98.91	43.40 (19.97)
FSMA (*n* = 38)	9 – 51.50	24.82 (10.91)
Adaptive level		
VABS ABC SS (*n* = 39)	44 – 83	61.23 (10.38)
VABS ABC AE (*n* = 39)	8.96 – 42.17	25.29 (8.54)

*PDD-NOS = Pervasive Developmental Disorder – Not Otherwise Specified

Table 3

Polychoric correlations among all YMQI items

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.00																			
2	0.11	1.00																		
3	0.27	0.06	1.00																	
4	0.19	0.21	0.57	1.00																
5	-0.03	0.01	0.07	0.25	1.00															
6	0.11	0.10	0.14	0.24	0.23	1.00														
7	0.40	0.14	0.40	0.40	0.24	0.02	1.00													
8	0.31	0.19	0.03	0.14	0.11	0.10	0.37	1.00												
9	0.29	0.16	0.22	0.36	0.16	0.22	0.22	0.38	1.00											
10	0.33	0.10	0.21	0.24	0.17	0.23	0.31	0.45	0.49	1.00										
11	0.37	0.12	0.13	0.27	0.02	0.34	0.29	0.48	0.54	0.51	1.00									
12	0.29	0.08	0.26	0.32	0.02	-0.09	0.35	0.41	0.43	0.25	0.33	1.00								
13	0.17	-0.05	0.14	0.25	0.11	0.07	0.18	0.30	0.43	0.34	0.29	0.39	1.00							
14	0.30	0.16	0.36	0.40	0.15	0.12	0.32	0.34	0.42	0.36	0.38	0.45	0.32	1.00						
15	0.11	0.08	0.32	0.23	0.15	0.14	0.23	0.15	0.28	0.12	0.14	0.23	0.17	0.74	1.00					
16	0.37	0.10	0.47	0.43	0.15	0.18	0.44	0.21	0.36	0.34	0.36	0.29	0.34	0.74	0.64	1.00				
17	0.36	0.10	0.45	0.56	0.17	0.29	0.53	0.24	0.38	0.32	0.37	0.29	0.29	0.67	0.54	0.75	1.00			
18	0.36	0.07	0.26	0.18	-0.06	0.07	0.33	0.35	0.24	0.24	0.26	0.48	0.17	0.58	0.39	0.51	0.62	1.00		
19	0.17	0.07	0.05	0.19	0.12	0.19	0.25	0.35	0.46	0.50	0.40	0.12	0.27	0.39	0.21	0.36	0.48	0.34	1.00	
20	0.16	0.13	0.03	0.14	0.09	0.09	0.14	0.40	0.46	0.48	0.31	0.19	0.37	0.29	0.10	0.27	0.30	0.17	0.49	1.00
21	0.52	0.00	0.28	0.38	0.07	0.11	0.43	0.28	0.32	0.30	0.46	0.48	0.22	0.34	0.12	0.38	0.45	0.40	0.29	0.23
22	0.67	0.00	0.36	0.44	0.14	0.18	0.56	0.27	0.32	0.42	0.45	0.36	0.31	0.35	0.13	0.49	0.61	0.47	0.38	0.25
23	0.03	0.27	-0.11	0.17	0.00	0.22	0.17	0.13	0.16	0.04	0.19	0.02	0.09	0.05	-0.05	0.11	0.12	0.08	0.19	0.05
24	0.10	0.13	0.16	0.22	0.11	0.16	0.18	0.34	0.13	0.26	0.20	-0.01	0.12	0.24	0.14	0.21	0.27	0.21	0.33	0.26
25	0.13	0.14	0.19	0.30	0.17	0.37	0.31	0.03	0.23	0.22	0.18	0.06	0.20	0.15	0.17	0.25	0.36	-0.05	0.23	0.19
26	0.18	0.26	0.22	0.29	0.22	0.41	0.37	0.07	0.13	0.17	0.20	-0.05	0.00	0.14	0.12	0.21	0.30	0.06	0.22	0.13
27	0.05	0.08	0.14	-0.20	-0.20	0.05	0.01	-0.07	-0.16	0.02	-0.04	-0.02	-0.08	0.01	-0.06	-0.02	0.00	0.03	0.01	-0.10
28	0.21	0.17	0.19	0.17	0.19	0.33	0.25	0.17	0.12	0.07	0.18	-0.04	-0.06	0.15	0.18	0.15	0.24	0.22	0.13	-0.01
29	0.34	-0.07	0.12	0.22	0.07	0.11	0.20	0.21	0.22	0.22	0.33	0.38	0.18	0.16	0.08	0.08	0.22	0.33	0.15	0.13
30	0.22	0.02	-0.03	-0.02	0.05	0.13	0.19	0.22	0.16	0.12	0.21	0.09	0.19	0.05	0.05	0.18	0.25	0.19	0.19	0.08
31	0.25	0.04	0.01	0.09	0.06	0.11	0.26	0.22	0.30	0.23	0.34	0.04	0.15	0.21	0.05	0.21	0.35	0.24	0.30	0.18

Item	21	22	23	24	25	26	27	28	29	30	31
21	1.00										
22	0.82	1.00									
23	0.16	0.21	1.00								
24	0.10	0.19	0.12	1.00							
25	0.12	0.21	0.17	0.06	1.00						
26	0.17	0.33	0.18	0.12	0.57	1.00					
27	0.11	0.09	0.16	-0.12	-0.04	0.06	1.00				
28	0.21	0.18	0.07	0.12	0.13	0.24	0.11	1.00			
29	0.56	0.55	0.05	0.08	0.01	0.08	0.21	0.14	1.00		
30	0.10	0.23	0.14	0.14	0.10	0.05	0.02	0.08	0.15	1.00	
31	0.21	0.34	0.14	0.21	-0.02	0.07	0.01	0.21	0.26	0.35	1.00

Table 4

Model Fit Statistics

Fit statistic	3-factor model	4-factor model	5-factor model
RMS	.08	.06	.05
RMSEA	.13	.12	.11
TLI	.54	.61	.64
BIC	473.97	69.18	-33.95

Table 5

Summary of Exploratory Factor Analysis Results for the YMQI Using Ordinary Least Squares Estimation – 4-Factor Model (N = 402)

	Factor Loadings				Communalities
Item	Factor 1	Factor 2	Factor 3	Factor 4	
15. Accessible materials	**0.898**	-0.054	-0.187	0.000	0.692
14. Clean plan and teaching goals	**0.843**	0.194	-0.028	-0.093	0.836
16. Length of inter-trail intervals	**0.741**	0.006	0.177	0.115	0.735
17. Suitable pace for the child	**0.608**	0.033	0.292	0.239	0.757
18. Intensive teaching	**0.506**	0.070	0.318	-0.224	0.501
3. Rapid reinforcer delivery	**0.435**	-0.189	0.235	0.180	0.316
4. Motivating reinforcers	**0.300**	0.047	0.214	0.299	0.361
9. Effectiveness of prompts	0.110	**0.690**	-0.059	0.051	0.527
8. Differential reinforcement	0.002	**0.689**	-0.013	-0.118	0.458
10. Fading and augmenting of prompts	-0.044	**0.638**	0.112	0.104	0.495
20. Evidence of skill acquisition	-0.027	**0.636**	-0.050	0.059	0.379
11. Lack of prompting errors	-0.018	**0.619**	0.155	0.064	0.503
19. Suitable task difficulty	0.085	**0.530**	0.058	0.164	0.424
13. Implementation of error correction	0.099	**0.414**	0.093	-0.045	0.259
24. Mixing tasks	0.087	**0.291**	-0.012	0.130	0.141
31. Use of prevention strategies	-0.012	**0.277**	0.212	0.025	0.176
2. Varying S^Ds	0.067	**0.216**	-0.185	**0.215**	0.102
30. Reinforcement of appropriate behaviour	-0.043	**0.199**	0.137	0.065	0.084
22. Maintenance of the child's focus	0.003	-0.002	**0.972**	0.098	0.995
21. On-task following requests	0.015	0.071	**0.809**	-0.086	0.699
1. Attending during S^Ds	0.030	0.077	**0.632**	-0.022	0.460
29. Result of problem behaviour	-0.100	0.113	**0.594**	-0.163	0.371
7. Sincere/motivating verbal reinforcement	0.193	0.048	**0.397**	0.234	0.393
27. Response generalization	-0.014	-0.157	0.177	-0.029	0.031
26. Teaching embedded in naturalistic activities	-0.009	0.016	0.120	**0.679**	0.511
25. Teaching away from the table	0.063	0.131	-0.070	**0.644**	0.459
6. Relation of reinforcer to the task	-0.002	0.202	-0.066	**0.513**	0.313
12. Follow through	0.254	0.306	0.233	**-0.338**	0.402
5. Varying reinforcers	0.072	0.087	-0.019	**0.288**	0.112
28. Flexible teaching	0.113	0.036	0.051	**0.248**	0.107
23. Varying teaching materials	-0.125	0.157	0.121	**0.225**	0.106

Table 6

Correlations Among Subscales

Subscale	1	2	3	4
1 (Pace and Organization)	1.00			
2 (Technical Correctness)	0.14	1.00		
3 (Engagement and Motivation)	0.19	0.15	1.00	
4 (Generalization)	-0.22	-0.10	-0.19	1.00

Table 7

Comparisons of the Fit of Linear and Quadratic Models for All Subscales

	Subscale 1		Subscale 2		Subscale 3		Subscale 4	
Model	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Linear	42.83	66.22	-187.87	-164.48	-36.47	-13.09	-196.68	-173.29
Quadratic	41.84	69.12	-187.31	-160.03	-41.73	-14.45	-196.04	-196.04
Likelihood ratio χ^2	2.99		1.44		7.26		1.36	
Likelihood ratio test *p*	0.08		0.23		0.007		0.24	

Table 8

Latent Growth Models with the Growth Factor of the Level of Subscale 1 at the Start of Treatment Regressed on Child Characteristics at the Start of IBI

Predictor	β	SE(β)	t	p
Autism Severity	0.00	0.00	0.94	0.36
Cognitive Level	-0.00	0.00	-0.18	0.86
Adaptive Level	-0.01	0.00	-0.70	0.49

Table 9

Latent Growth Models with the Growth Factor of the Level of Subscale 2 at the Start of Treatment Regressed on Child Characteristics at the Start of IBI

Predictor	β	SE(β)	t	p
Autism Severity	-0.00	0.00	-0.30	0.76
Cognitive Level	-0.00	0.00	0.43	0.67
Adaptive Level	-0.00	0.00	-0.16	0.87

Table 10

Latent Growth Models with the Growth Factor of the Level of Subscale 3 at the Start of Treatment Regressed on Child Characteristics at the Start of IBI

Predictor	β	SE(β)	*t*	*p*
Autism Severity	-0.01	0.00	-2.18	0.04
Cognitive Level	0.00	0.00	1.27	0.21
Adaptive Level	0.00	0.00	0.19	0.85

Table 11

Latent Growth Models with the Growth Factor of the Level of Subscale 4 at the Start of Treatment Regressed on Child Characteristics at the Start of IBI

Predictor	β	SE(β)	*t*	*p*
Autism Severity	0.00	0.00	0.30	0.77
Cognitive Level	-0.00	0.00	-0.58	0.57
Adaptive Level	0.00	0.00	0.62	0.54

Table 12

Spearman Correlations of Subscales at the Start of IBI and the Change in Child Characteristics from Time 1 to Time 2

	Subscale 1	Subscale 2	Subscale 3	Subscale 4
Change in Autism Severity	.004	-.653**	-.311	-.486*
Change in Cognitive Level	-.109	.213	.242	.472*
Change in Adaptive Level	.076	-.043	.045	.073

* $p < .05$ ** $p < .01$.

Table 13

Linear Regression Models with the Slope of Subscale 1 and the Level of the Same Child Characteristic at Time 1 Regressed on Child Characteristics at Time 2 (Outcome)

Predictor	β	SE(β)	t	p
Autism Severity	18.83	73.91	0.26	.80
Cognitive Level	-21.60	173.60	-0.12	.90
Adaptive Level	-76.90	176.17	-0.44	.67

Table 14

Linear Regression Models with the Slope of Subscale 2 and the Level of the Same Child Characteristic at Time 1 Regressed on Child Characteristics at Time 2 (Outcome)

Predictor	β	SE(β)	t	p
Autism Severity	234.82	119.91	1.96	0.06
Cognitive Level	-62.03	320.50	-0.19	0.85
Adaptive Level	-276.17	324.24	-0.85	0.40

Table 15

Linear Regression Models with the Slope of Subscale 3 and the Level of the Same Child Characteristic at Time 1 Regressed on Child Characteristics at Time 2 (Outcome)

Predictor	β	SE(β)	t	p
Autism Severity	144.37	80.89	1.79	0.08
Cognitive Level	-191.57	209.33	-0.92	0.37
Adaptive Level	-92.12	217.49	-0.42	0.67

Table 16

Linear Regression Models with the Slope of Subscale 4 and the Level of the Same Child Characteristic at Time 1 Regressed on Child Characteristics at Time 2 (Outcome)

Predictor	β	SE(β)	t	p
Autism Severity	144.37	80.89	1.79	0.08
Cognitive Level	-191.57	209.33	-0.92	0.37
Adaptive Level	-92.12	217.49	-0.42	0.57

Table 17

Linear Regression Models with the Slope of Subscale 1 and the Level of the Same Child Characteristic at Time 1 Regressed on the Change in Child Characteristics during IBI

Predictor	β	SE(β)	t	p
Autism Severity	111.23	67.57	1.65	0.11
Cognitive Level	-126.72	168.74	-0.75	0.46
Adaptive Level	-213.30	179.03	-1.19	0.24

Table 18

Linear Regression Models with the Slope of Subscale 2 and the Level of the Same Child Characteristic at Time 1 Regressed on the Change in Child Characteristics during IBI

Predictor	β	SE(β)	t	p
Autism Severity	315.19	124.64	2.53	0.02
Cognitive Level	-117.83	332.22	-0.36	0.73
Adaptive Level	-371.43	349.73	-1.06	0.30

Table 19

Linear Regression Models with the Slope of Subscale 3 and the Level of the Same Child Characteristic at Time 1 Regressed on the Change in Child Characteristics during IBI

Predictor	β	SE(β)	t	p
Autism Severity	200.50	83.76	2.39	0.02
Cognitive Level	-165.72	218.74	-0.76	0.45
Adaptive Level	-39.11	237.03	-0.17	0.87

Table 20

Linear Regression Models with the Slope of Subscale 4 and the Level of the Same Child Characteristic at Time 1 Regressed on the Change in Child Characteristics during IBI

Predictor	β	SE(β)	t	p
Autism Severity	200.50	83.76	2.39	0.02
Cognitive Level	-165.72	218.74	-0.76	0.45
Adaptive Level	-39.11	237.03	-0.16	0.87

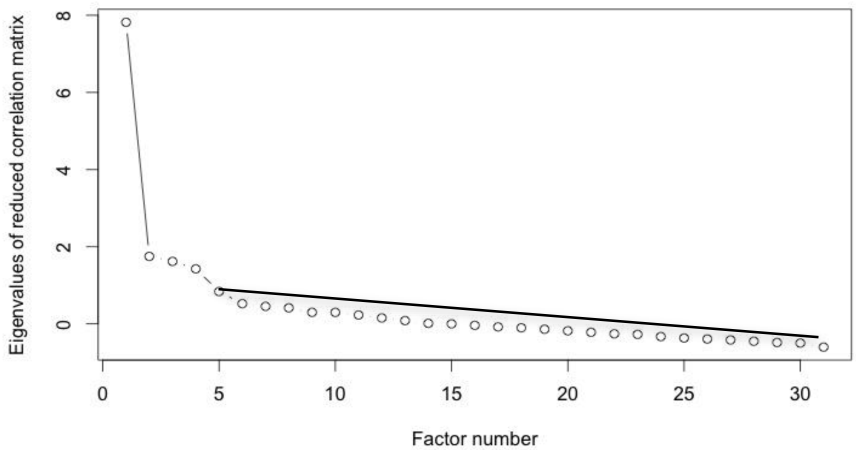

Figure 1. Scree plot suggesting a 4-factor structure of the YMQI.

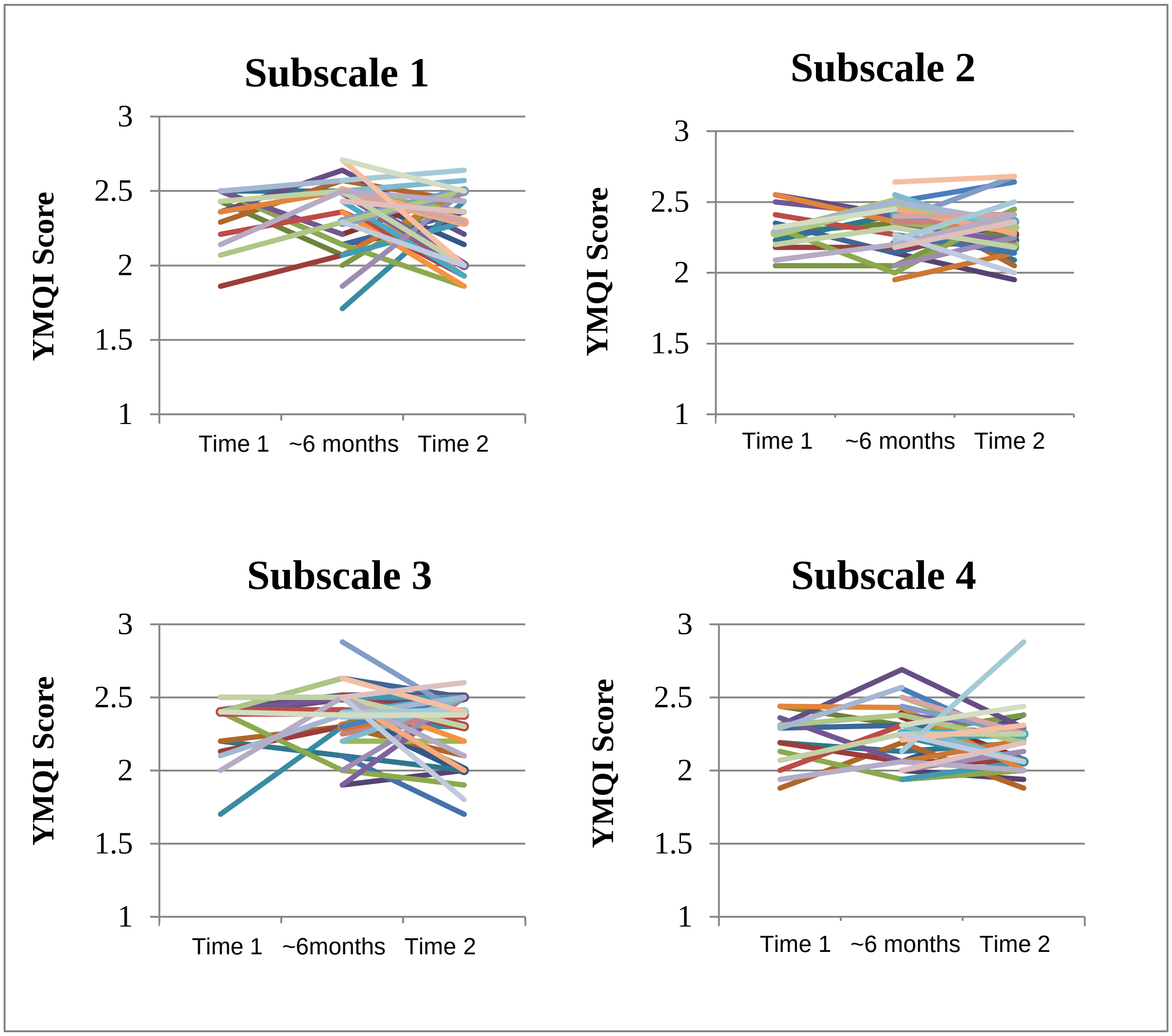

Figure 2. Individual children's YMQI subscales across three time points during IBI.

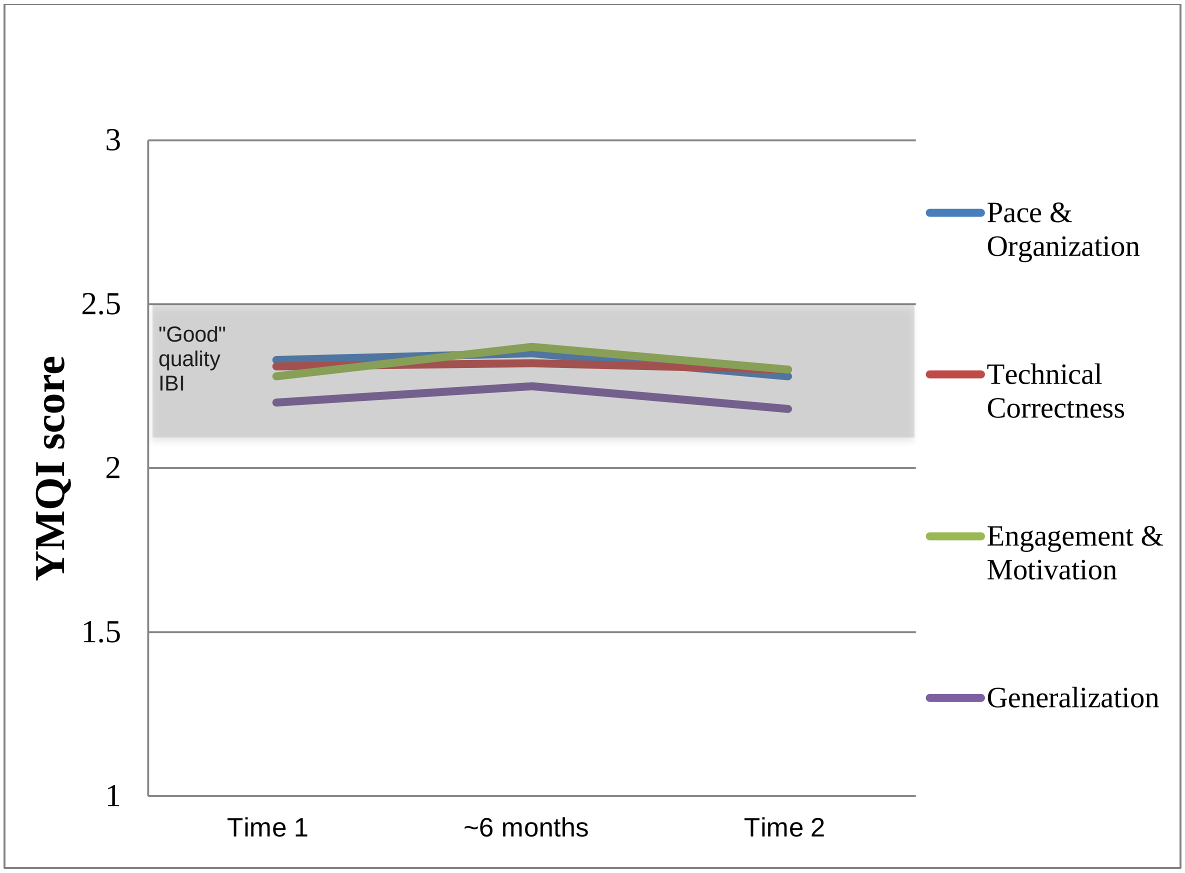

Figure 3. Means of YMQI subscales across approximately one year in IBI.

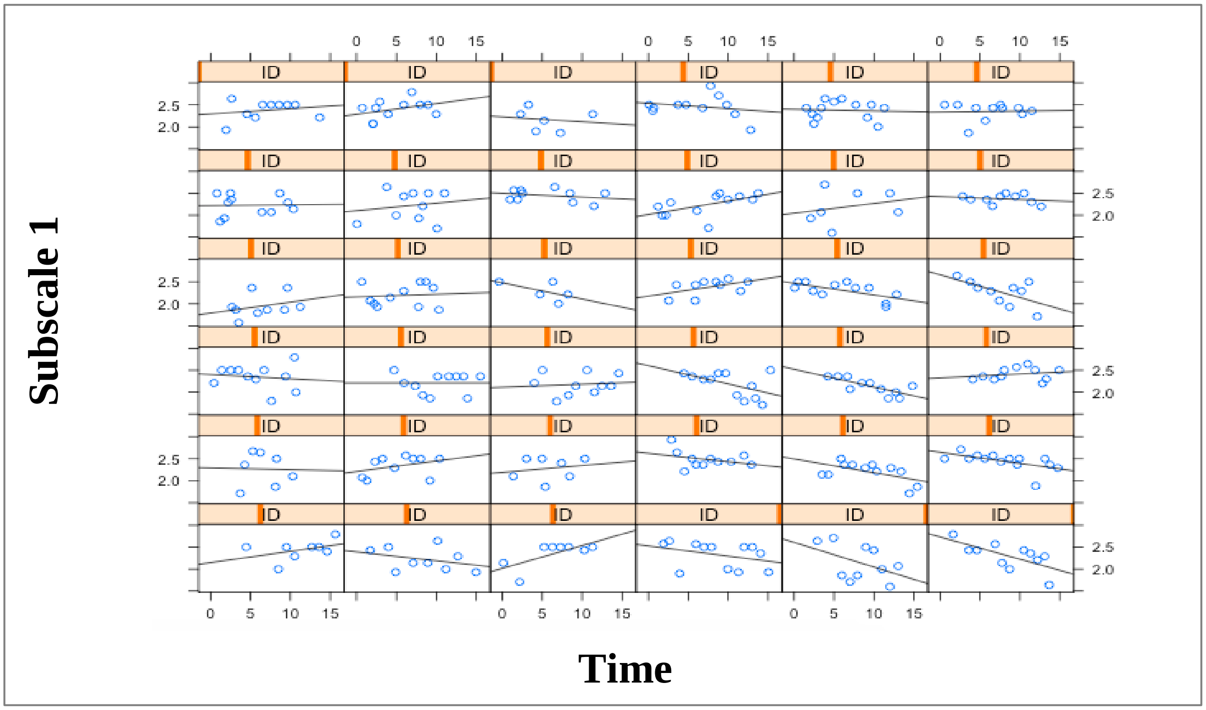

Figure 4. Individual children's data across all time points for subscale 1 (Pace and Organization).

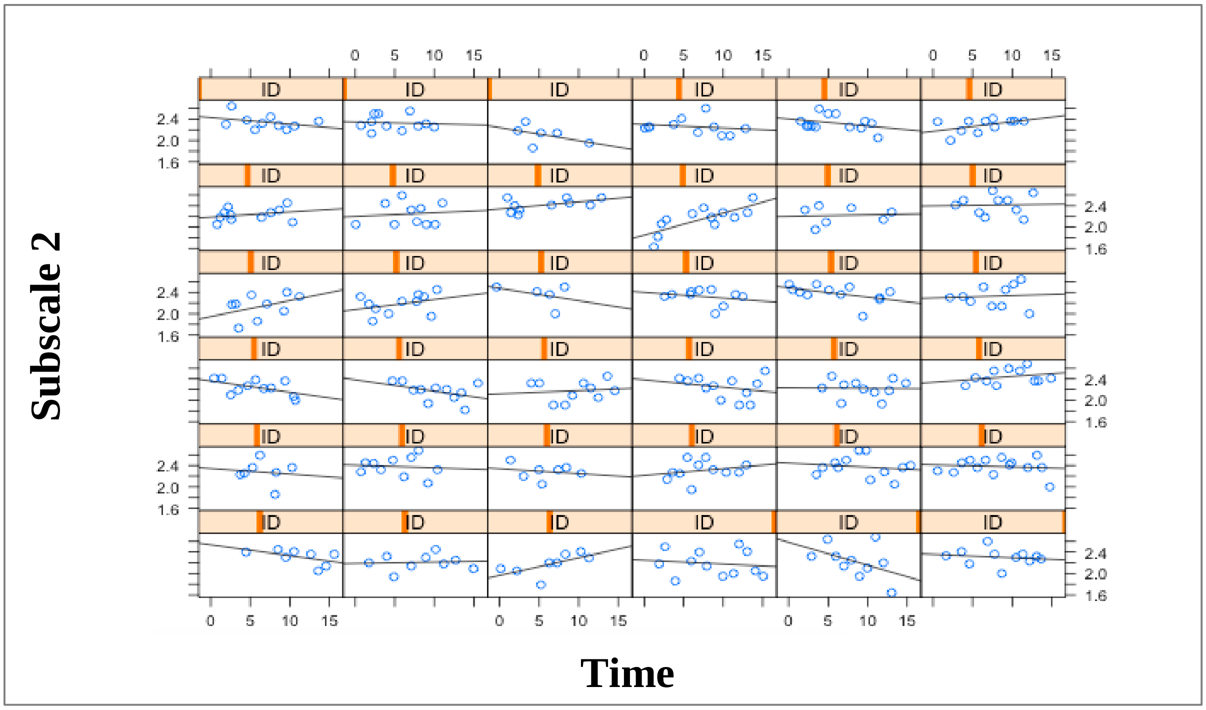

Figure 5. Individual children's data across all time points for subscale 2 (Technical Correctness).

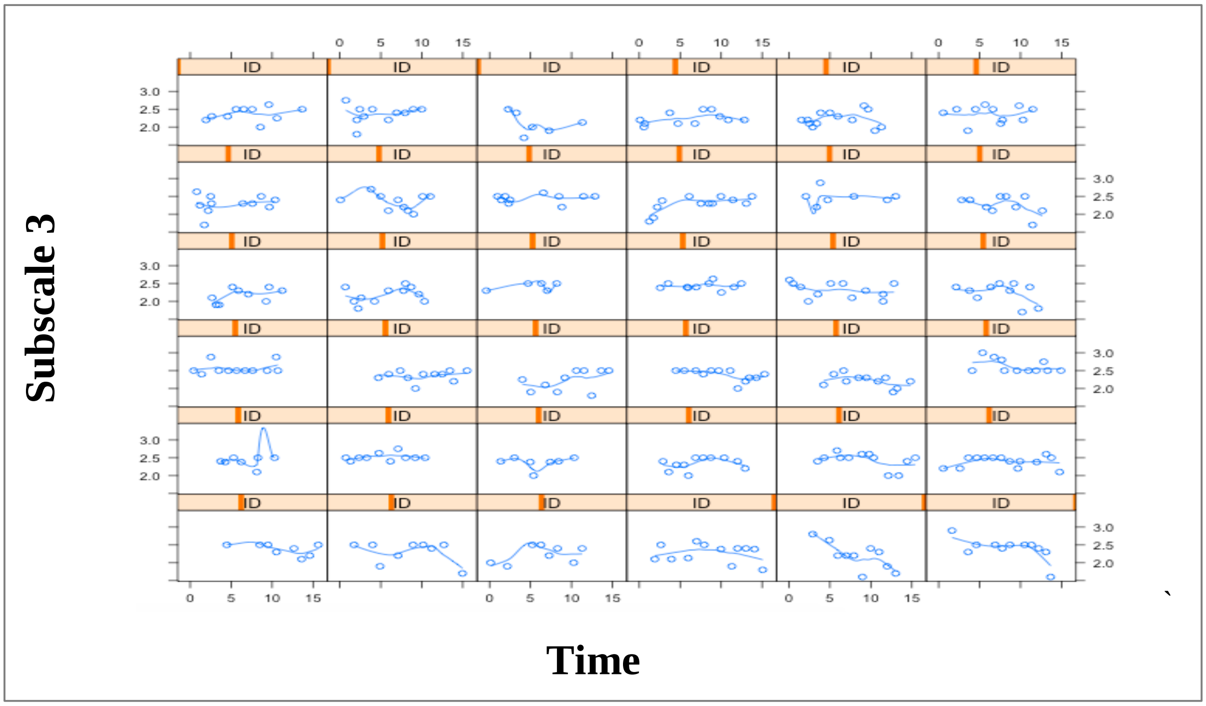

Figure 6. Individual children's data across all time points for subscale 3 (Engagement and Motivation).

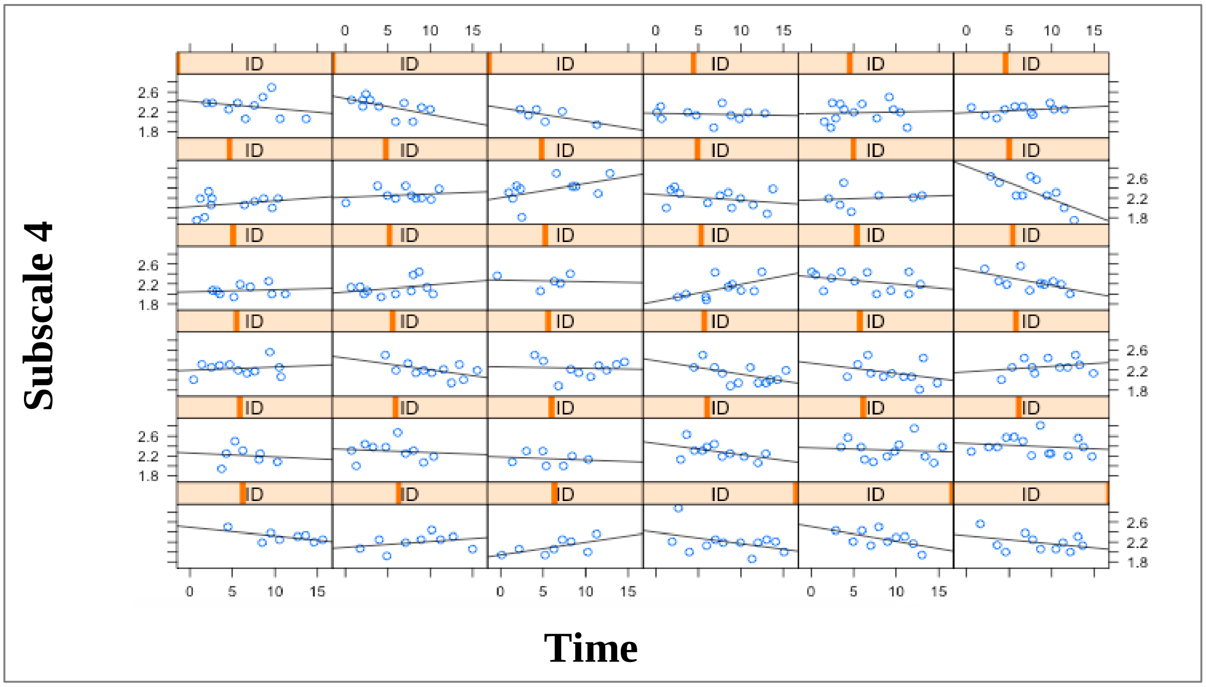

Figure 7. Individual children's data across all time points for subscale 4 (Generalization).