Research in Autism Spectrum Disorders 7 (2013) 721–740

Contents lists available at SciVerse ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

An experimental analysis of the effects of therapeutic horseback riding on the behavior of children with autism☆

Sarah R. Jenkins, Florence D. DiGennaro Reed *

Department of Applied Behavioral Science, University of Kansas, United States

ARTICLE INFO

Article history:
Received 19 December 2012
Received in revised form 22 February 2013
Accepted 24 February 2013

Keywords:
Therapeutic horseback riding
Single-case design
Alternative treatments

ABSTRACT

The current study experimentally evaluated the effects of therapeutic horseback riding on the behavior of children with autism using a multiple baseline across participants design and a waitlist control group for comparison purposes. Participants were observed weekly in an after-school program during four center-based activities and during therapeutic horseback riding lessons. We also conducted intermittent probes of behavior at home. Therapeutic horseback riding did not produce systematic changes in affect, responding to others' initiations, spontaneous initiations, off-task behavior, compliance, problem behavior, or performance on two standardized measures. Three of four participants' posture improved during therapeutic horseback riding.

The prevalence of Autism Spectrum Disorder (ASD) has risen nearly 600% in the past two decades with current estimates suggesting that 1 in 88 children have a diagnosis (Centers for Disease Control and Prevention, 2011). Reports project that the number of children with ASD will exceed the number of children with cancer, juvenile diabetes, and pediatric AIDS combined (Autism Speaks, n.d.c; Centers for Disease Control and Prevention, 2011; National Standards Project, 2009). The nation spends $137 billion per year on costs associated with ASD (Autism Speaks, n.d.b) while families spend a staggering $1.2 million caring for an individual with ASD across their lifetime (Autism Speaks, n.d.a). The collective individual and societal costs for treatment indicate a need to select and adopt interventions that have been shown to be effective in high quality, experimental research (National Standards Project, 2009). However, a myriad of treatment options makes it challenging to select an appropriate, empirically supported course of treatment. For example, Romanczyk, Gillis, White, and DiGennaro (2008) identified 414 different interventions for children with ASD. Fortunately, standards for evaluating the efficacy of interventions have been developed to help guide treatment decisions.

Evidence-based practice (EBP) refers to treatment strategies and clinical techniques supported by high-quality experimental research, expertise from clinicians, and client preferences (Kazdin, 2008). A necessary—though not sufficient—component of EBP is reliance on empirically validated interventions (Kazdin, 2008). The criteria used for empirical validation differs as a function of the research methodology adopted by single-subject researchers (i.e., behavior analysts) and more traditional psychological researchers who use large sample sizes, which contributes to the lack of consensus across disciplines regarding the criteria that constitute EBP. Additionally, there is variability among the terminology used (e.g., empirically validated therapy, scientifically supported treatment) which may be confusing to consumers, caregivers, and

☆ This study was conducted by the first author in partial fulfillment of the requirements for the M.A. in behavior analysis at the University of Kansas. We would like to extend our appreciation to all of the undergraduate research assistants who collected data and assisted in other ways on this project. We very much thank you for your consistent dedication and hard work.

* Corresponding author at: Department of Applied Behavioral Science, University of Kansas, 4001 Dole Human Development Center, 1000 Sunnyside Avenue, Lawrence, KS 66045-7555, United States. Tel.: +1 785 864 0521.
E-mail address: fdreed@ku.edu (F.D. DiGennaro Reed).

1750-9467/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rasd.2013.02.008

clinicians (DiGennaro Reed & Reed, 2008). Adding to this confusion, particularly for caregivers, is that descriptions of various treatment options may suggest empirical support where none exists (e.g., complementary and alternative medicines; CAM).

CAM includes procedures, interventions, and/or treatments that are not considered conventional for a particular treatment need (Umbarger, 2007). These interventions can be used in conjunction with conventional medicine (complementary) or in lieu of common practices (alternative). Types of CAM include therapies targeting the body and mind, biological approaches, manipulation of the body, approaches addressing energy, and a whole system approach, which integrates all of the approaches described (Atkins, Angkustsiri, & Hansen, 2010). The use of CAM has increased, in part, due to greater accessibility through media outlets, personal contacts, and practitioners using alternative medicines (Nickel & King Gerlach, 2001). Recent estimates suggest that, among families with children with disabilities, CAM use is highest for families with children with ASD (Atkins et al., 2010). Animal-assisted therapy (AAT) is one type of CAM and, as the name implies, incorporates the use of animals into occupational, physical, or other therapeutic procedures.

Therapeutic horseback riding (THR) is a type of AAT that teaches horsemanship skills such as holding a horse's reins appropriately, controlling the horse with voice commands, and other basic riding skills (Bracher, 2000; Drnach, O'Brien, & Kreger, 2010). Goals of therapy include improving balance, posture, gross and fine motor skills, and communication (Bertoti, 1988; Snider, Korner-Bitensky, Kammann, Warner, & Saleh, 2007).

Past research has reported benefits in varying domains of development when THR is used as an intervention for children with ASD (Bass, Duchowny, & Llabre 2009; Gabriels et al., 2012; Kern et al., 2011; Nelson et al., 2011). For example, Bass et al. (2009) found improvements in social functioning, measured by standardized questionnaires, for an experimental group (19 children diagnosed with autism) after a 12-week THR program, compared to a waitlist control group (15 children with ASD) who did not participate in THR. In another study, Kern et al. (2011) reported improvements in the severity of symptoms commonly associated with ASD for 20 participants who completed a 6-month THR riding program using the Childhood Autism Rating Scale (Schopler, Reichler, & Renner, 1994). The riding program was evaluated within a group design with repeated measurement, using both clinician and parent-rated assessments and questionnaires. Using standardized behavioral and physical assessments by individuals who were not blind to the study, Gabriels et al. (2012) reported improvements in self-regulation behaviors, expressive language, and motor skills for a group of 26 children with ASD who participated in a 10-week THR program. A waitlist control group (16 children with ASD) did not demonstrate the same changes. These three studies that evaluated the effects of THR lacked direct observation and measurement of behavior. Instead, they relied on surveys and standardized behavior assessments. Only one study to date used a single-case research design and direct measurement of behavior in order to evaluate the effects of THR. Nelson et al. (2011) used a single-case reversal design and reported improvements in social behavior for three children with ASD who were exposed to THR for a total of 2.5 h. Baseline consisted of noncontingent access to activities; however, the horses were positioned on the opposite end of the arena (i.e., participants did not ride the horse). The treatment phase consisted of similar materials and activities as baseline, while the participants rode the horse and practiced various skills. Although the authors attribute increases in performance to THR, changes may be due to other confounding variables. For example, programmed reinforcement was provided for social behavior during baseline and THR sessions. Increases in socialization may be a function of reinforcement, rather than the intervention. In addition, measurement of social behaviors aggregated both independent and prompted responses and it is unclear which form of social behaviors increased.

Despite the positive benefits reported by these researchers, reviews conducted by several entities (e.g., Association for Science in Autism Treatment, n.d.; National Standards Project, 2009; Umbarger, 2007) indicate that THR does not meet the criteria for an EBP. Parents of children with ASD may still be inclined to select THR as a treatment because descriptions of its effects can be readily found on the internet (Wong & Smith, 2006). Online resources, such as parent blogs, might suggest positive benefits of THR in the absence of a methodologically rigorous evaluation. Moreover, parents may not have the scientific background and training to evaluate scientific rigor and identify poorly designed studies (Kay & Vyse, 2008). In a field that supports and "demands" the use of evidence-based interventions it would be beneficial to scientifically document the effects of this treatment option. The current study used a single-case experimental design to evaluate the effects of THR on the behavior of children with ASD. We selected target behaviors that were similar to the outcomes evaluated in prior studies, but could be measured using direct observation and recording techniques. We also included pre- and post-tests of two standardized measures including the Child Behavior Checklist (Achenbach & Rescorla, 2001) and the Bruininks–Oseretsky Test of Motor Proficiency, second edition (Bruininks & Bruininks, 2005).

1. Method

1.1. Participants

Seven children (six boys, one girl) with autism between 6 and 14 years of age ($M = 9.5$ years) participated in the study. Two of the participants were identical twin boys. Participants were diagnosed before the start of the study by an independent, licensed professional. To be included in the study, participants could not have prior exposure to THR or hippotherapy and were required to live within 30 miles of the primary research site, a horse arena located in the Midwest. Participants were recruited via flyers and e-mails disseminated to local public schools. Parents of the participants did not already have their child enrolled in a THR program. We divided participants into a treatment (Seth, Selina, Frank, and Milo) and a waitlist control group (Ivan, Denis, and Edmund). Participants assigned to the latter group did not receive THR until

Table 1
Participant descriptions.

Name	Age	Diagnosis	VABS-II scores	Description of skills
Seth	6	ASD	ABC = 58 MBC = 17	Vocal communication skills, tact colors and shapes, read sight words, and count to 100
Selina	13	ASD, Verbal and Motor Apraxia	ABC = 57 MBC = 16	Vocal communication skills; tact colors, shapes, and the alphabet; read sight words
Frank	6	ASD	ABC = 36 MBC = 19	Nonvocal; match stimuli; receptively identify numbers, letters, shapes, and colors
Milo	14	Tuberous sclerosis, ASD	ABC = 42 MBC = 19	Nonvocal; match stimuli; requires assistance with writing; receptively identify some colors
Ivan	14	Tuberous sclerosis, ASD	ABC = 40 MBC = 22	Nonvocal; match identical stimuli; requires hand-over-hand assistance when writing
Denis	6	ASD	ABC = 58 MBC = 19	Vocal communication skills; expressively identify some colors, shapes, and letters; and count to 50
Edmund	8	ASD	ABC = 64 MBC = 20	Vocal communication skills, tact colors, add and subtract numbers, read

Note: ASD = autism spectrum disorder, VABS-II = Vineland Adaptive Behavior Scales, second edition, ABC = adaptive behavior composite, MBC = maladaptive behavior composite.

after the completion of the study. All participants received a scholarship for THR so that families did not incur any costs for their child's participation. See Table 1 for a more detailed description of each participant.

1.2. Setting

Participants were observed during an after-school program held in a reserved area at the riding facility where THR was provided. The after-school program observations were conducted weekly during four center-based activities including an academic task, art, games, and snack. Activities took place on a tabletop containing relevant materials for the academic, art, and game centers (e.g., crayons, puzzles, flashcards). During snack, participants sat on a couch to eat food items provided by parents. Dyads rotated through the centers, which were led by a research assistant. Two activities, games and snack, were located in separate rooms measuring 3.66 m × 3.54 m and 3.96 m × 4.75 m, respectively. Art and the academic centers were located in the same room with tables placed 0.61 m apart. All participants were also observed in their home during typical routines on three occasions (i.e., single session probes across phases) to assess generalization of behaviors. Homes were located within 30 miles of the horse facility and included single-family homes and townhomes. The presence of other individuals (e.g., caregivers, siblings, and friends) varied across families and observations. Treatment group participants were also observed during weekly THR sessions held in an arena measuring 24.38 m × 30.48 m. THR sessions were held immediately following the after-school program observations. One to four participants were present during THR sessions atop horses. In addition, two side-walkers and one horse leader were also present for each participant, while a certified instructor was positioned in the middle of the arena and delivered the therapy.

1.3. Dependent variables and response measurement

Primary data collection occurred once a week for 10 min for each of four center-based activities (40 min total) and during 10-min intermittent home probe observations. We collected data during weekly THR sessions for treatment group participants; however, these data were not subject to experimental analysis because it was not possible to adopt a multiple baseline design across the THR sessions. This was due, in part, to dependent variables that could not measured during a true baseline, such as posture and commands given to the horse, because they were based on the participants' behavior while riding the horse. Without a true baseline, this aspect of the study lacks experimental control.

Several behaviors and dimensions of behavior were recorded for each participant. The primary dependent variables were selected based on variables measured in previous research (that could be measured using direct observation) and claims reported on parent blogs about the effects of THR. Table 2 provides detailed information about the dependent variables including operational definitions, dimensions, calculations, and setting(s).

Two of the dependent variables were measured using percent of opportunity (i.e., compliance and responses to initiations). The number of opportunities was identical during each session. Research assistants were instructed to deliver particular instructions or initiations via a written checklist, which presented different opportunities for paired participants.

1.3.1. Affect (happiness and unhappiness)

Happiness was defined as "any facial expression or vocalization typically considered an indicator of happiness among people without disabilities" (Green & Reid, 1999, p. 284). Unhappiness was defined as "any facial expression or vocalization typically considered an indicator of unhappiness among people without disabilities" (Green & Reid, 1999, p. 284).

Table 2
Description of dependent variables.

Dependent variable	Definition	Data collection technique and dimension	Calculation	Setting
Affect: happiness	"Any facial expression or vocalization typically considered an indicator of happiness among people without disabilities" (Green & Reid, 1999, p. 284) (e.g., smiling, laughing, yelling while smiling)	10 s MTS % of intervals	(# Intervals with happiness/total number of intervals) × 100	H, Art and Academic
Affect: unhappiness	"Any facial expression or vocalization typically considered an indicator of unhappiness among people without disabilities" (Green & Reid, 1999, p. 284) (e.g., frowning, grimacing, crying, scowling, yelling without smiling)	10 s MTS % of intervals	(# Intervals with unhappiness/total number of intervals) × 100	H, Art and Academic
Spontaneous initiations	Any language used before a prompt or model was provided (Matson et al., 1990). A new initiation was scored when a pause occurred for 10 s during which the child did not use any type of language	Rate/min	Frequency/10 min	H, C, A
Responses to initiations	Any contextually appropriate vocalization, picture exchange, use of augmentative device, sign language, or other form of communication within 3 s of another's initiation (adapted from Shafer et al., 1984)	% of opportunity	(# of responses/total # of opportunities) × 100	H, C, A
Off-task behavior	Motor behaviors or verbalizations that are not permitted or are unrelated to the current task (e.g., not seated at the center, manipulating materials in a way that is not appropriate for the task) (adapted from DiGennaro et al., 2007, p. 449)	10 s MTS % of intervals	(# intervals with off-task bx/total number of intervals) × 100	Art and Academic
Compliance	Following a direction within 10 s of its presentation (adapted from Wilder et al., 2006)	% of opportunity	(# of directions followed/ total # of opportunities) × 100	H, Snack and Games, A
Problem behavior	Aggression (e.g., hitting, slapping, kicking, biting, pushing) directed toward another individual; pica (eating non-food items); stereotypy (e.g., hand flapping, body rocking, finger posturing, non-contextual vocalizations); screaming or other vocalizations not appropriate for the setting; property destruction (e.g., inappropriate tearing, throwing, ripping materials); and any other disruptions not appropriate for the setting (e.g., jumping on furniture)	10 s MTS % of intervals	(# intervals with problem bx/total number of intervals) × 100	H, C, A
Commands to direct the horse	Tapping the horse on the neck or vocalizations (e.g., walk-on, whoa, or trot) delivered by the participant in order to get the horse to walk, turn, or stop	Rate/min	Frequency/10 min	A
Posture	Appropriate posture was defined as sitting upright with the back parallel to the wall and buttocks in the saddle. Deviations from sitting upright by more than 45° were considered examples of inappropriate posture	10 s MTS % of intervals	(# of intervals with appropriate posture/total number of intervals) × 100	A

Note: H = home visit probes, C = all center-based activities, A = arena, Bx = behavior, MTS = momentary time sampling.

1.3.2. Responses to initiations

Responses to initiations were defined as any contextually appropriate vocalization, picture exchange, use of augmentative device, sign language, or other form of communication within 3 s of another's initiation (adapted from Shafer, Egel, & Neef, 1984).

1.3.3. Spontaneous initiations

Spontaneous initiations were defined as any language (e.g., vocal, sign language) used before a prompt or model was provided (Matson, Sevin, Fridley, & Love, 1990).

1.3.4. Off-task behavior

Off-task behavior was defined as any motor behaviors or verbalizations that are not permitted or are unrelated to the current task (adapted from DiGennaro, Martens, & Kleinmann, 2007).

1.3.5. Compliance

Compliance was defined as following a direction within 10 s of its presentation (adapted from Wilder, Atwell, & Wine, 2006).

1.3.6. Problem behavior

Problem behavior was defined broadly as any disruptive behavior not appropriate for the setting (e.g., aggression, property destruction, pica, stereotypy).

1.3.7. Commands to direct the horse

Commands were defined as tapping the horse on the neck or vocalizations (e.g., "walk-on," "whoa," or "trot") delivered by the participant in order to get the horse to walk, turn, or stop.

1.3.8. Posture

Appropriate posture was defined as sitting upright with the back parallel to the wall and buttocks in the saddle.

1.3.9. Standard scores

Two standardized measures were administered before and after THR. The Child Behavior Checklist (CBCL/6–18: Achenbach & Rescorla, 2001) was used to assess problem behaviors and competencies using rating scales given to teachers (teacher rating form; TRF) and parents and was administered before and after the 9-week therapy program. The CBCL and TRF are comprised of profiles and scales which measure competence, adaptive functioning, internalizing and externalizing problems, and syndromes (e.g., somatic complaints, thought problems). For this study, only externalizing, internalizing, and total problems scores were reported. *T*-scores were used to determine the range of behaviors displayed within each behavior category. For the internalizing, externalizing, and total problems categories, *T*-scores above 63 indicate clinical ranges and *T*-scores between 60 and 63 indicate borderline ranges, while *T*-scores below 60 are considered to be within normal ranges. Internal consistency coefficients range from 0.63 to 0.79 for competence scales, 0.78–0.97 for problem scales, and 0.72–0.91 for the DSM-oriented scales; these coefficients overall range from moderately high to high internal consistency.

In addition, we attempted to administer the Bruininks–Oseretsky Test of Motor Proficiency, second edition (Bruininks & Bruininks, 2005) to further evaluate changes in posture; however, many of the participants did not have the comprehension skills necessary to comply with the assessment tasks and their data could not be scored. For this reason, these data are not included in this study.

1.4. Experimental design and procedures

A multiple baseline design across participants was used to evaluate the effects of THR on the behavior of treatment group participants in the after-school program and at home. The analysis consisted of two phases: (a) baseline, and (b) THR.

1.4.1. Baseline

During baseline, participants were observed during a weekly after-school program and at home before receiving THR. Participants rotated through the four centers in pairs. During the center-based activities in the after-school program, participants did not receive programmed reinforcement for appropriate behaviors (e.g., engagement, on-task behavior) or contingent on particular performance. There were also no programmed consequences for problem behavior. During home visit probes, parents interacted with their children as they typically would while at home. The next phase was introduced when all dependent variables were stable or displaying a trend in the direction opposite to that anticipated during the intervention.

1.4.2. THR

The purpose of this phase was to evaluate the effects of THR on participant behavior. Four participants received THR; the remaining participants were assigned to the control group. Weekly 60-min therapy sessions were conducted during an established 9-week THR program. The riding program was accredited by the Professional Association of Therapeutic Horsemanship (PATH) International, a nonprofit organization, established in 1969 (PATH International, n.d.). The same instructor taught all but one lesson for all participants throughout the study. Preparation for THR involved creating lesson plans based on each rider's skill level and acquisition of target horsemanship skills. At the time of the study, the instructor had 5 years of experience with THR, four of which she practiced with the PATH International credential. Two volunteers (side-walkers) walked alongside the horse during THR for each participant to ensure their safety during the riding lesson. The side-walkers also provided verbal prompts or physical guidance when riders were instructed to deliver a command to the horse or to comply with the activities. A third volunteer, a horse leader, also provided assistance during the session by guiding the horse around the arena if the rider was not able to deliver commands to control the horse. The number of volunteers required to walk with participants was determined by the instructor who administered a pre-screening horse evaluation before THR. All four participants who received THR required two side-walkers and one horse leader. During this phase, observations during the center-based activities and home visit probes continued similar to baseline.

1.5. Interobserver agreement and procedural fidelity

1.5.1. Interobserver agreement (IOA)

Observers independently collected data on all dependent variables during at least 30% of sessions during the after-school program and home visit probes and at least 30% of THR sessions for all participants. IOA for interval recording (i.e., affect,

off-task behavior, and problem behavior) was calculated using the interval-by-interval method, by dividing the number of intervals with agreement on the occurrence or nonoccurrence of the dependent variables by the total number of intervals, multiplied by 100. IOA for rate-based behaviors (i.e., spontaneous initiations and commands given to the horse) was calculated using the total count approach, in which the smaller rate was divided by the larger rate, multiplied by 100. IOA for behaviors recorded per opportunity (i.e., compliance and responses to initiations) was calculated using an adaptation of the trial-by-trial approach, in which the number of opportunities with agreement on occurrence or nonoccurrence of behavior was divided by the total number of opportunities, multiplied by 100. During 60% of sessions in baseline, IOA averaged 96% across all dependent variables, participants, and settings (range, 92–100%). Agreement across all dependent variables for 51% of the home and center activity observations within the THR phase averaged 97% (range, 86.8–100%). During 36% of sessions of THR, agreement averaged 98.4% (range, 89–100%) across all dependent variables.

1.5.2. Procedural fidelity

Undergraduate research assistants completed a protocol checklist during the games and snack centers to ensure they provided opportunities for participants to display skills and behaviors we measured (e.g., responses to other's initiations and compliance). Although the materials varied weekly (i.e., parents sent in different food each week, activities during the games center changed each week), the number of opportunities to comply or respond to initiations was held constant from week to week. These sessions were video-recorded and scored by an independent observer during 92% of sessions to confirm fidelity of implementation. Fidelity was calculated by dividing the sum of the number of steps followed correctly (e.g., delivering the correct trials and circling the appropriate answer choice) divided by the total number of steps in the protocol, multiplied by 100. Procedural fidelity averaged 96.6% (range, 89–100%). Procedural fidelity was also recorded for 30% of therapy sessions for Seth, Selina, and Frank and 43% of therapy sessions for Milo. The THR instructor provided a lesson plan for the session and an independent observer recorded whether each step was implemented as indicated and in the correct order. To calculate procedural fidelity, the sum of the correct steps implemented by the instructor was divided by the total number of steps, multiplied by 100. Procedural fidelity averaged 86.7% (range, 60–100%).

2. Results

2.1. Center activities and home visits

2.1.1. Affect

The percentage of intervals in which happiness and unhappiness occurred was recorded during two 10-min center activities (art and academic) and during 10-min home visit probes. Data for the treatment and control group are depicted in Figs. 1 and 2, respectively.

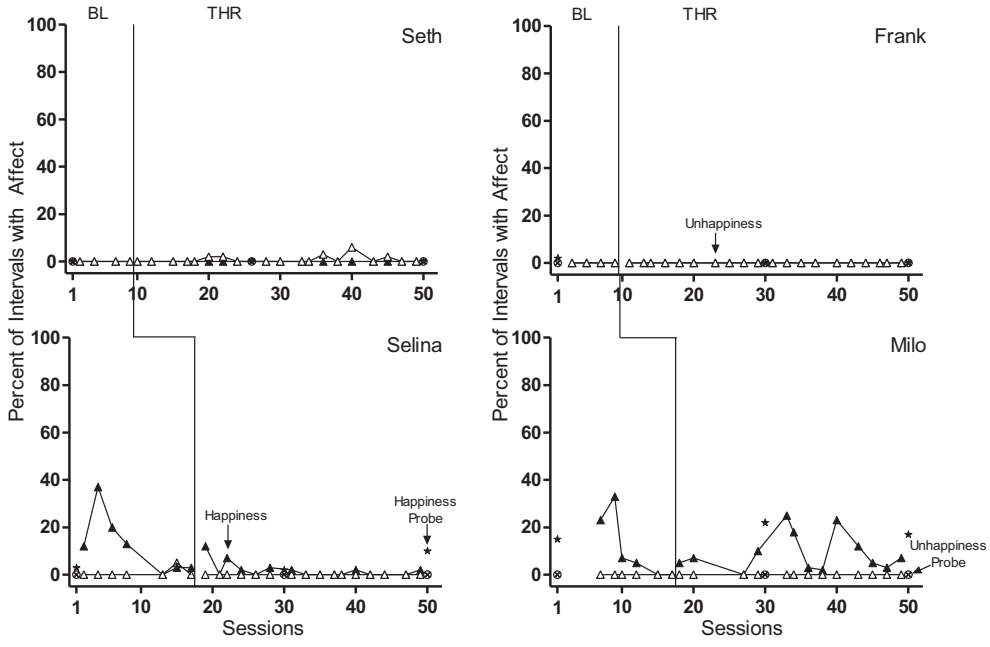

Fig. 1. Percentage of intervals with indices of happiness and unhappiness for the treatment group. Closed triangles indicate happiness during center-based activities, stars indicate happiness during home visit probes, open triangles denote unhappiness during center-based activities, and circles with an × indicate unhappiness during home visit probes.

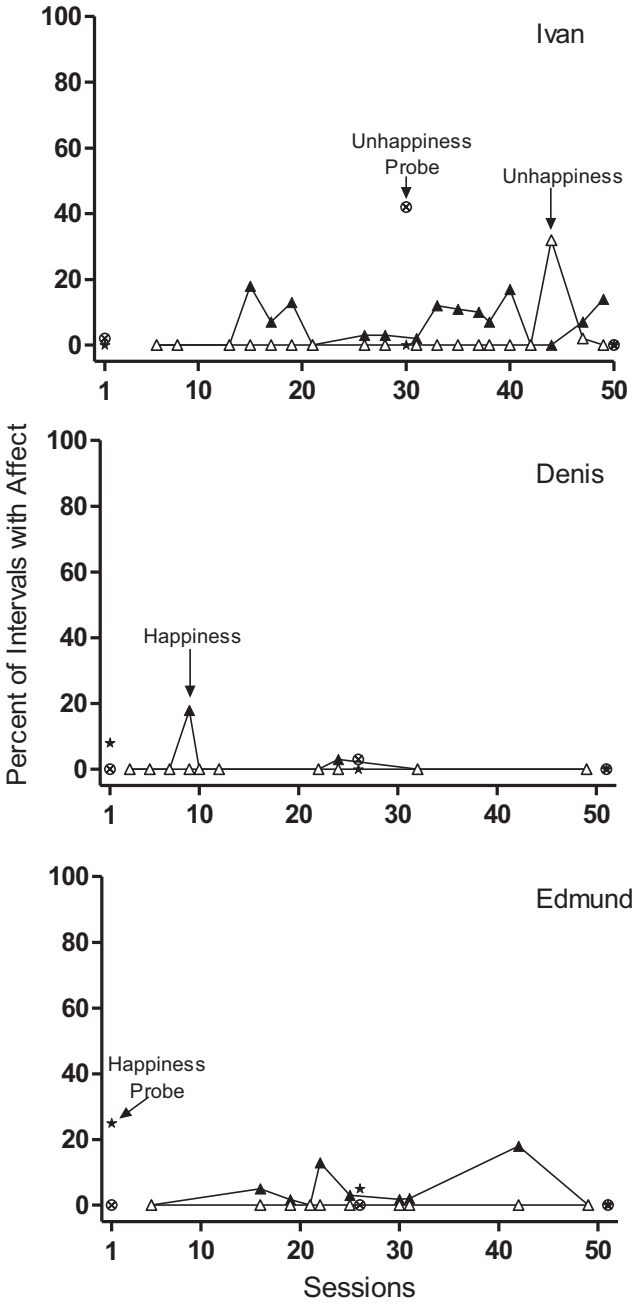

Fig. 2. Percentage of intervals with indices of happiness and unhappiness for the control group. Closed triangles indicate happiness during center-based activities, stars indicate happiness during home visit probes, open triangles denote unhappiness during center-based activities, and circles with an × indicate unhappiness during home visit probes.

2.1.1.1. Treatment group. Seth's happiness ($M = 0\%$) and unhappiness ($M = 0\%$) during baseline in the center activities were low and stable. He displayed similar levels of happiness and unhappiness during the home visit probe (0% for both). When THR was introduced, happiness and unhappiness were also low and stable and remained unchanged though there was increased variability for unhappiness (happiness, $M = 0\%$; unhappiness, $M = 0.79\%$; range, 0–6%). During the two home visit probes during the THR phase, happiness ($M = 0\%$) and unhappiness ($M = 0\%$) remained unchanged and were low and stable. Selina's happiness ($M = 12.57\%$; range, 0–37%) and unhappiness ($M = 0.71\%$; range, 0–5%) were also low in baseline during center activities. Her percentage of happiness shows a decreasing trend during baseline and was somewhat variable; unhappiness was stable. During the baseline home visit probe, happiness and unhappiness was observed during 3% and 0% of intervals, respectively. There was no change in happiness and unhappiness once THR was introduced for center activities (happiness, $M = 1.88\%$; range, 0–12%; unhappiness, $M = 0\%$) or home visit probes (happiness $M = 6\%$; range, 2–10%; unhappiness $M = 0\%$). Frank's baseline percentage of happiness and unhappiness during center activities were at zero levels. During the home visit probe in baseline, happiness and unhappiness occurred during .67% of intervals and 0% of intervals, respectively. During the THR phase, happiness and unhappiness remained at 0% of intervals for both center activities and

home visit probes. Milo's baseline happiness (M = 11.33%; range, 0–33%) during center activities was low and somewhat variable with a sharp decreasing trend, while unhappiness was low and stable (M = 0%). He displayed 15% and 0% of intervals of happiness and unhappiness, respectively, during the home visit probe. Happiness was somewhat variable in the center activities during the THR phase (M = 9.23%; range, 0–25%); unhappiness continued to occur at zero levels during center activities. Milo's level of happiness (M = 19.5%; range, 17–22%) was also somewhat variable during the home visit probes, while unhappiness (M = 0%) was similar to baseline levels. In sum, the treatment group data indicate that THR did not produce changes in affect during center activities or home visit probes.

2.1.1.2. Control group. During center activities Ivan's level of happiness was somewhat variable (M = 6.5%; range, 0–18%). Unhappiness was low and stable with a brief increase for one session (M = 1.79%; range, 0–32%). Happiness observed during home visit probes was variable (M = 14.7%; range, 0–42%), while unhappiness was at zero levels for all home visit probes. Baseline happiness for Denis during center activities was low and stable with a one-session increase (M = 2.1%; range, 0–18%). Unhappiness was at zero levels. During the home visit probes, happiness (M = 2.67%; range, 0–8%) and unhappiness (M = 1%; range, 0–3%) were both low and stable. Happiness for Edmund was somewhat variable and low during center activities (M = 4.45%; range, 0–18%). This pattern was also observed during home visit probes (M = 10%; range, 0–25%). Unhappiness, however, was displayed at zero levels during the center activities and home visit probes.

2.1.2. Responses to initiations

The percentage of responses to initiations was recorded during two 10-min center activities (games and snack) as well as 10-min home visit probes. Data for the treatment and control group are depicted in Figs. 3 and 4, respectively.

2.1.2.1. Treatment group. Seth's percentage of responses to initiations during center activities was low and at zero levels except for one session during baseline (M = 12.5%; range, 0–50%). The percentage of responses made during the baseline home visit probe was 100% with two opportunities to respond. In the THR phase, responses to initiations during center activities were low and at zero levels except for two sessions during which the percentage of responses increased (M = 4.15; range, 0–50%). During the home visit probes, the percentage of responses was 33%; however, this percentage only represents responses made during one of the home visits; there were no opportunities to respond during the last home visit probe during the THR phase. Selina's percentage of responses to initiations during baseline in center activities was high, variable, and increasing (M = 84%; range, 50–100%). The percentage of responses during the baseline home visit probe was 70% with 10 opportunities to respond. After the introduction of THR, responses to initiations during center activities decreased initially and then became high and slightly variable during centers activities (M = 88.33%; range, 25–100%). Performance in the home visit probes during THR was variable (M = 50%; range, 0–100%). During the first probe, her performance was 100% out of six opportunities, but decreased to 0% in the final observation during which there was one opportunity to respond. The percentage of responses to initiations for Frank during baseline in center activities was low and stable (M = 0%). This was also the case during the baseline home visit probe, despite the presentation of 15 opportunities to respond. During the THR phase,

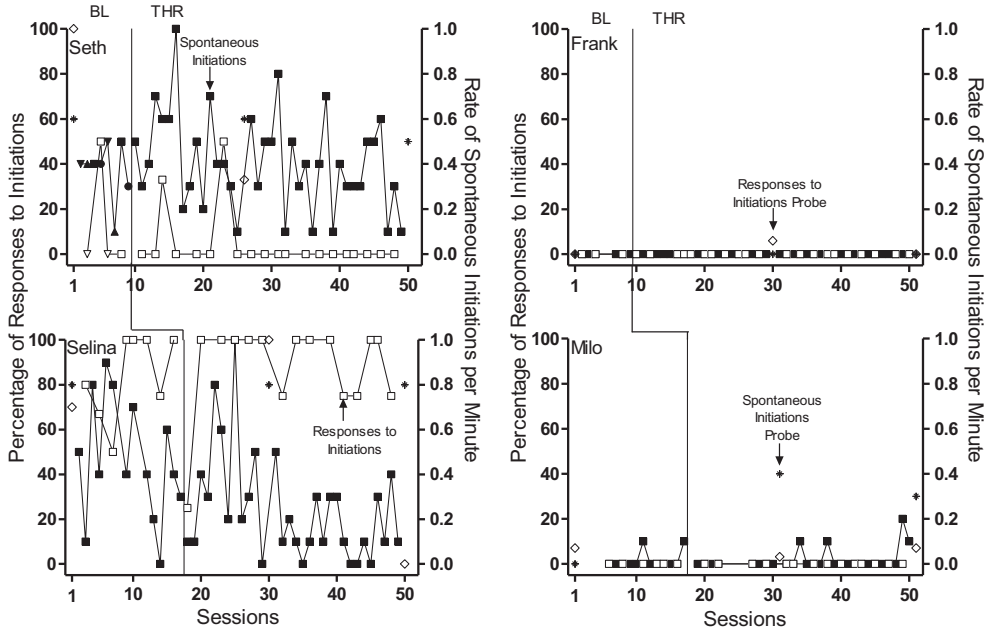

Fig. 3. Rate of spontaneous initiations and percentage of responses to others' initiations for the treatment group. Closed squares indicate spontaneous initiations during center-based activities, stars indicate spontaneous initiations during home visit probes, open squares indicate responses to others' initiations during center-based activities, and open diamonds indicate responses to others' initiations during home visit probes.

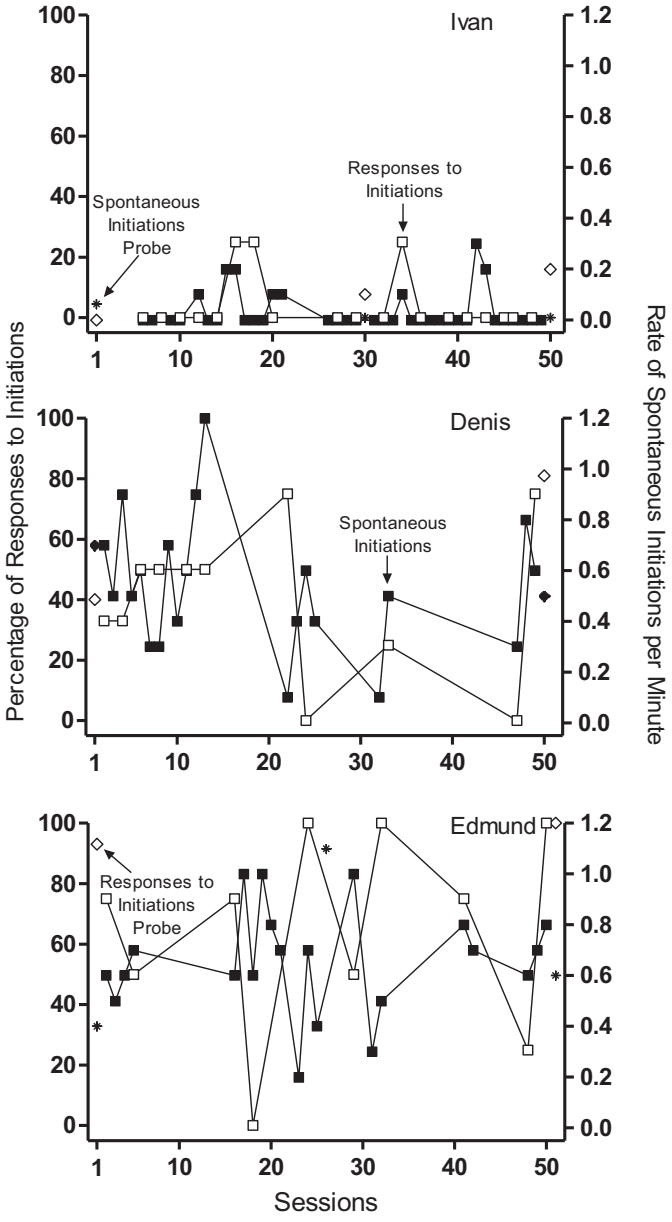

Fig. 4. Rate of spontaneous initiations and percentage of responses to others' initiations for the treatment group. Closed squares indicate spontaneous initiations during center-based activities, stars indicate spontaneous initiations during home visit probes, open squares indicate responses to others' initiations during center-based activities, and open diamonds indicate responses to others' initiations during home visit probes.

Frank's performance remained low and stable during the center activities ($M = 0\%$) and the home visit probes ($M = 3\%$; range, 0–6%). Twenty-one opportunities were presented during the first home observation in this phase, while five opportunities were presented during the second home observation. Milo did not make any responses in baseline during center activities. During the baseline home visit probe, responses occurred in 7% of the 29 presented opportunities. During the THR phase, responses did not occur during center activities. An average of 5.1% responses (range, 3.2–7%) occurred during the home visit probes. Thirty-two opportunities were presented during the first home visit probe and 14 opportunities were presented during the second home visit probe. In sum, these data indicate that THR had no effect on participants' responses to initiations.

2.1.2.2. Control group. Ivan's baseline percentage of responses to initiations was low and somewhat stable during center activities ($M = 3.95\%$; range, 0–25%) and low during the home visit probes ($M = 1.5\%$ range, 0–4.5%). The number of opportunities to respond during home visit probes ranged from 15 to 22. The baseline percentage of responses to initiations for Denis had an increasing trend initially, but became variable during center activities ($M = 40.1\%$; range, 0–75%) and the home visit probes ($M = 61.7\%$; range 40–81%). The number of opportunities to respond during the home visit probes ranged from 11 to 21. Edmund's percentages had high variability during center activities in baseline ($M = 65\%$; range, 0–100%). During the home visit probes, his percentages of responses to initiations were high ($M = 96.5\%$; range, 93–100%) and stable.

The opportunities presented to respond during these visits ranged from 0 to 14 for two of the home visits. No opportunities were presented during one of the home visit probes; therefore, those data are not represented in Fig. 4.

2.1.3. Spontaneous initiations

The rate of spontaneous initiations was recorded during all four 10-min center activities as well as during 10-min home visit probes. Data for the treatment and control group are depicted in Figs. 3 and 4, respectively.

2.1.3.1. Treatment group. Spontaneous initiations in baseline during center activities for Seth was stable early in the phase, but increased in variability over time ($M = 0.38$; range, 0.1–0.5). During the baseline home visit probe, the rate of spontaneous initiations was 0.6. In the THR phase during center activities, the rate of spontaneous initiations increased initially and then became highly variable ($M = 0.41$; range, 0.1–1). During the home visit probes, spontaneous initiations were stable ($M = 0.55$; range, 0.5–0.6). Selina's rate of spontaneous initiations in baseline during center activities had a decreasing trend with variability ($M = 0.46$; range, 0–0.9). During the baseline home visit probe, the rate was 0.8. In the THR phase, rates were variable with a general decreasing trend across the phase ($M = 0.25$; range, 0–1). During the THR home visit probes, the rate of spontaneous initiations was stable and averaged 0.8. During baseline, Frank's rate of spontaneous initiations in center activities was low and stable; spontaneous initiations did not occur during center activities or the home visit probe. This was also the case in the THR phase during center activities and the subsequent home visit probes. During baseline, Milo's rate of spontaneous initiations during center activities was low and somewhat stable ($M = 0.02$; range, 0–0.1). Spontaneous initiations did not occur during the baseline home visit probe. During the THR phase, the rate remained low and somewhat stable during center activities ($M = 0.02$; range, 0–0.2). During the home visit probes, spontaneous initiations were stable ($M = 0.35$; range, 0.3–0.4). Overall data for the treatment group indicate that THR did not produce changes in the rate of spontaneous initiations during center activities or home visit probes.

2.1.3.2. Control group. Ivan's rate of spontaneous initiations in baseline was low and somewhat variable during center activities ($M = 0.04$; range, 0–0.3). During the home visit probes, spontaneous initiations were low with an increasing trend ($M = 0.1$; range, 0–0.2). Denis had highly variable spontaneous initiations in baseline during center activities ($M = 0.54$; range, 0.1–1.2), while the rate during home visit probes was stable ($M = 0.63$; range, 0.5–0.7). The rate of spontaneous initiations for Edmund was also variable during center activities ($M = 0.66$; range, 0.2–1). Spontaneous initiations during the home visit probes were variable ($M = 0.7$, range, 0.4–1.1).

2.1.4. Off-task

The percentage of intervals in which off-task behavior occurred was recorded during two center activities (art and academic). Data for the treatment and control group are depicted in Figs. 5 and 6, respectively.

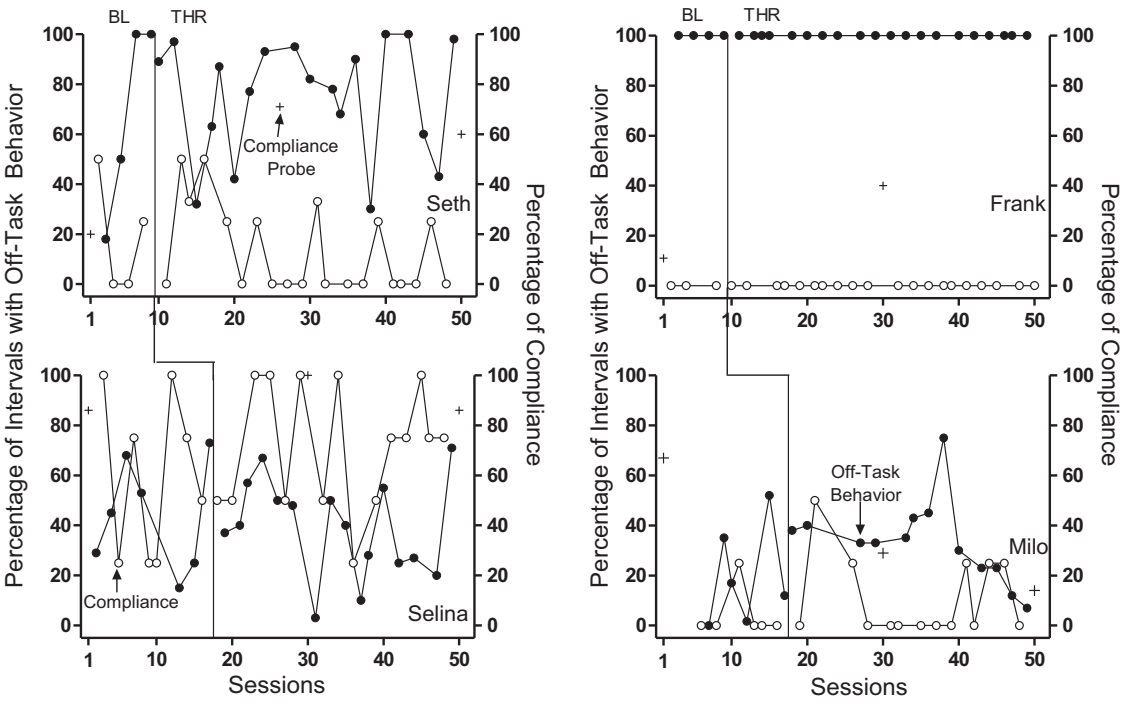

Fig. 5. Percentage of intervals with off-task behavior and the percentage of compliance for the treatment group. Closed circles indicate off-task behavior during center-based activities, closed circles indicate compliance during center-based activities, and pluses indicate compliance during home visit probes.

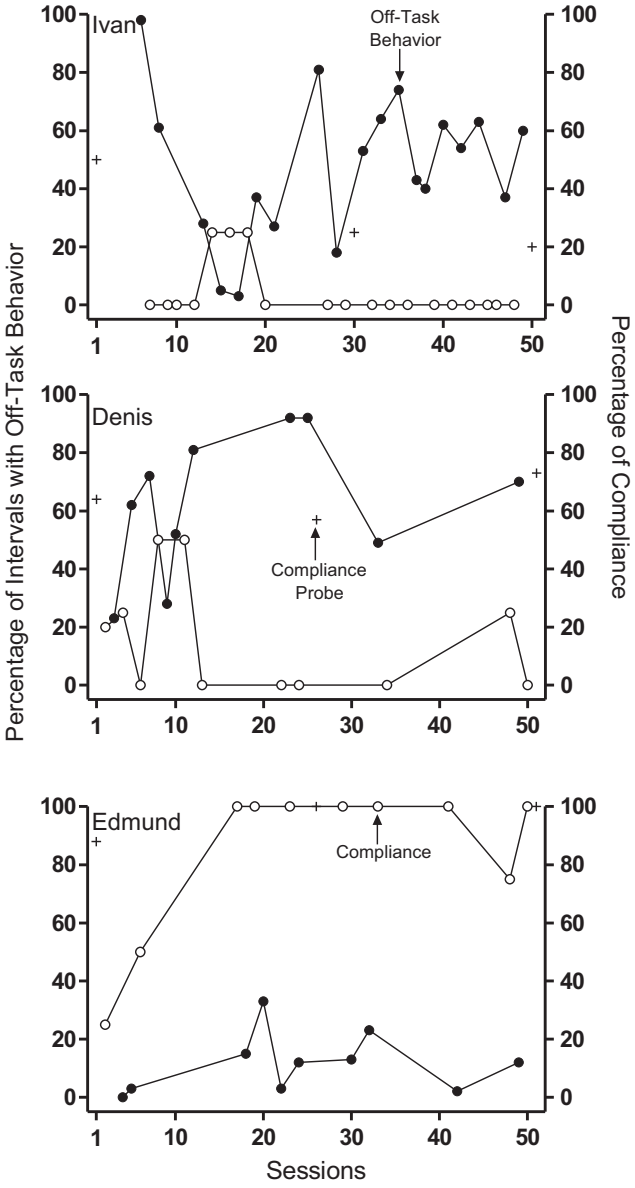

Fig. 6. Percentage of intervals with off-task behavior and the percentage of compliance for the control group. Closed circles indicate off-task behavior during center-based activities, closed circles indicate compliance during center-based activities, and pluses indicate compliance during home visit probes.

2.1.4.1. Treatment group. Seth's off-task behavior showed an increasing trend during baseline but stabilized at 100% of intervals at the end of the phase ($M = 67\%$; range, 18–100%). During the THR phase, the percentages were generally high with variability noted ($M = 74.95\%$; range, 30–100%). Selina's off-task behavior in baseline was variable with an increasing trend toward the end of the phase ($M = 44\%$; range, 15–73%). Off-task behavior remained unchanged during the THR phase and showed similar variability ($M = 39.3\%$; range, 3–71%). Frank's percentages in baseline were high and stable ($M = 100\%$). This pattern continued in the THR phase; Frank engaged in off-task behavior during 100% of intervals. Milo's off-task behavior in baseline was variable ($M = 19.62\%$; range, 0–52%). After the introduction of THR, Milo's off-task behavior increased and remained stable for about seven sessions with a steep increase in off-task behavior during the eighth session ($M = 33.62\%$; range, 7–75%). The last five sessions within this phase, however, showed a decreasing trend. In sum, data for this group indicate that THR did not produce clinically significant changes in off-task behavior.

2.1.4.2. Control group. Ivan's ($M = 47.79\%$; range, 3–98%) and Denis' ($M = 62.1\%$; range, 23–92%) percentage of off-task behavior in baseline during center activities was highly variable. Edmund's off-task behavior was low and variable ($M = 11.6\%$; range, 0–33%).

2.1.5. Compliance

The percentage of compliance (unprompted) to an adult direction was recorded during two 10-min center activities (games and snack) as well as 10-min home visit probes. Data for the treatment and control group are depicted in Figs. 5 and 6, respectively.

2.1.5.1. Treatment group. Compliance in baseline for Seth was generally low, but variable during the center activities (M = 18.75%; range, 0–50%). Seth complied one time out of five opportunities during the baseline home visit probe. During the THR phase, compliance was in the same range as baseline and showed similar variability (M = 13.3%; range, 0–50%). For the home visit probes, Seth complied with 71% of the seven directions that were presented during the first probe and he complied with 60% of the five directions presented in the second probe. Selina's percentages in baseline were variable (M = 59.38%; range, 25–100%). Compliance during the baseline home visit probe was 86% (seven directions were presented). Her range of compliance in the THR phase remained unchanged and was variable (M = 71.67%; range, of 25–100%). However, stability was observed by the end of this phase. During the first home visit probe in the THR phase, Selina complied with both directions presented to her. During the second probe, she complied with 86% of the seven directions that were presented. Frank's percentages during both baseline and treatment were low and stable; he did not comply with any directions during center activities. During the home visit probe in baseline, Frank complied with 11% of nine directions presented. For the first home visit probe in the THR phase, Frank complied with 33% of six directions. He did not independently comply with any of the 10 directions presented during the second home visit probe in this phase. Milo's baseline compliance was low and stable with one session during which compliance increased during center activities (M = 4.17%; range, 0–25%). During the home visit probe, he complied with 67% of the nine directions presented. In the THR phase, compliance during center activities was generally low with variability (M = 10.71%; range, 0–50%). During the home visit probes, Milo independently complied with 29% of the 14 directions presented during the first home visit probe in this phase and 14% of the 14 directions presented in the second home visit probe. In sum, these findings suggest that THR does not impact levels of compliance during center activities or home visit probes.

2.1.5.2. Control group. Ivan's compliance was stable and low during baseline in the center activities (M = 3.95%; range, 0–25%). During the home visit probes he complied with directions an average of 31.67% (range, 20–50%) without a prompt. The number of directions presented during these probes ranged from 5 to 22. Denis' baseline percentages during center activities were somewhat variable (M = 15.45%; range, 0–50%). Compliance during the home visit probes averaged 64.67% (range, 57–73%), with 7–17 opportunities to comply. The baseline percentages during center activities for Edmund show an increasing trend, with increased stability over time (M = 85%; range, 25–100%). During the home visit probes, compliance averaged 96% (range, 88–100%). Three to 16 directions were presented.

2.1.6. Problem behavior

The percentage of intervals in which problem behavior occurred was recorded during all four 10-min center activities as well as during 10-min home visit probes. Data for the treatment and control group are depicted in Figs. 7 and 8, respectively.

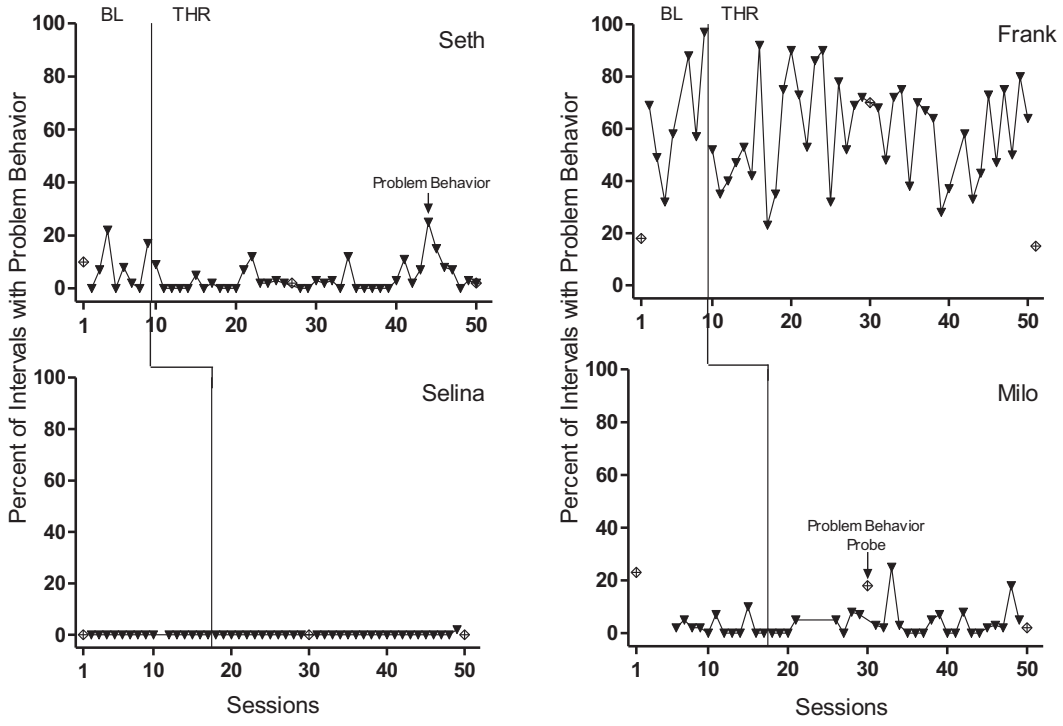

Fig. 7. Percentage of intervals with problem behavior for the treatment group. Closed upside-down triangles indicate problem behavior during center-based activities and open diamonds with a plus indicate problem behavior during home visit probes.

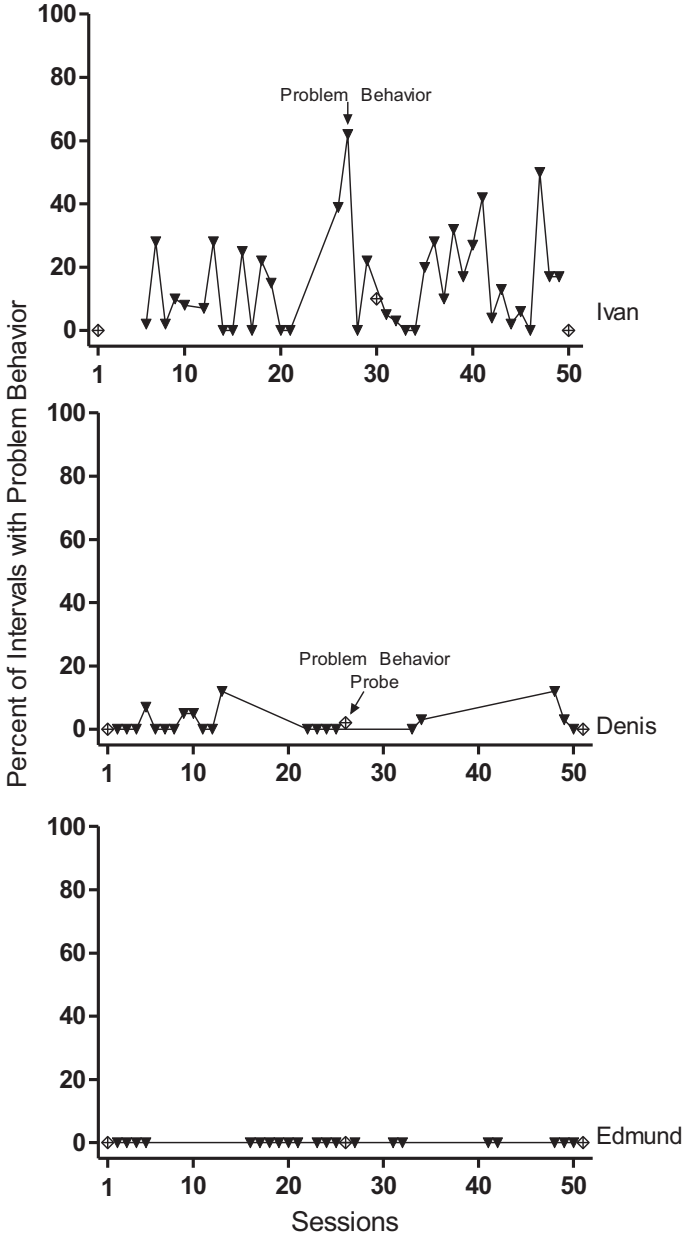

Fig. 8. Percentage of intervals with problem behavior for the control group. Closed upside-down triangles indicate problem behavior during center-based activities and open diamonds with a plus indicate problem behavior during home visit probes.

2.1.6.1. Treatment group. Seth's baseline percentage of intervals with problem behavior during center activities was somewhat variable and low (M = 7%; range, 0–22%). Problem behavior during the baseline home visit probe occurred during 10% of intervals. During the THR phase, the low and somewhat variable trend continued during center activities (M = 3.68%; range, 0–25%). During the home visit probes, problem behavior was low and stable (M = 2%). The percentage of intervals with problem behavior during both center activities and the home visit probe in baseline was zero for Selina. During the THR phase, problem behavior averaged 0.06% during center activities and 0% during the home visit probes. Frank's baseline percentage of intervals during which problem behavior occurred during center activities was variable with an increasing trend (M = 64.29%; range, 32–97%). During the home visit probe, problem behavior occurred during 18% of intervals. In the THR phase, problem behavior during center activities was in the similar range as baseline and showed similar variability (M = 58.44%; range, 23–92%). During the home visit probes, problem behavior occurred less, on average, than during center activities in this phase (M = 42.5%; range, 15–70%). Milo's baseline percentages were low and stable during center activities (M = 2.33%; range, 0–10%). For the home visit probe, problem behavior occurred during 23% of intervals. In the THR phase, his problem behavior was low and somewhat variable during center activities (M = 4%; range, 0–25%). Problem behavior during the home visit probes was lower, on average, compared to the baseline home visit probe (M = 10%; range, of 2–18%). Overall, data for the treatment group indicate that THR did not have an effect on the occurrence of problem behavior during center activities or home visit probes.

2.1.6.2. Control group. Ivan's baseline percentage of intervals with problem behavior during center activities was variable (M = 14.82%; range, of 0–62%). Problem behavior during home visit probes was low (M = 3.33%; range, 0–10%). Denis' percentages in baseline during center activities were low and somewhat stable (M = 2.24%; range, 0–12%). During the home visit probes, problem behavior was low and stable (M = 67%; range, of 0–2%). Edmund did not engage in problem behavior in the center activities or home visit probes.

2.2. THR

2.2.1. Posture

The percentage of intervals in which participants displayed appropriate posture while mounted on the horse was recorded for 10 min during each weekly lesson. These data are depicted in Fig. 9. Seth's percentage of intervals with appropriate posture was variable with stability observed during the final three lessons (M = 60.5%; range, 35–77%). Selina's percentages were high and stable (M = 94.25%; range, 90–98%). Frank's appropriate posture showed an increasing trend during THR (M = 64.5%; range, 0–92%). Milo's percentages were generally high and stable with a slight increase at the end of the study (M = 71.86%; range, 64–82%). These data suggest that participants' posture improved during THR.

2.2.2. Responses to initiations

The percentage of responses to another's initiation was captured using a voice recorder, in which a 10-min time sample of each THR session was scored for each participant. These data are depicted in Fig. 10. The percentage of Seth's (M = 13.89%; range, 0–50%) and Selina's (M = 70.83%; range, 0–100%) responses to initiations was variable Frank and Milo did not respond to any initiations made by others across all sessions. In sum, these data suggest that responses to initiations remain unchanged throughout THR.

2.2.3. Spontaneous initiations

The rate of spontaneous initiations was captured each week via a voice recorder, in which a 10 min time sample of each THR session was scored for each participant. These data are depicted in Fig. 10. Seth's rate had a decreasing trend (M = 0.58; range, 0.2–1). Selina's rate of spontaneous initiations was variable (M = 0.55; range, 0.1–1.1). Frank did not make any spontaneous initiations during THR. Milo had a low and stable rate (M = 0.01; range, 0–0.1). In sum, there were no changes in the rate of spontaneous initiations for three of four participants. One participant demonstrated decreases in the rate of spontaneous initiations during THR.

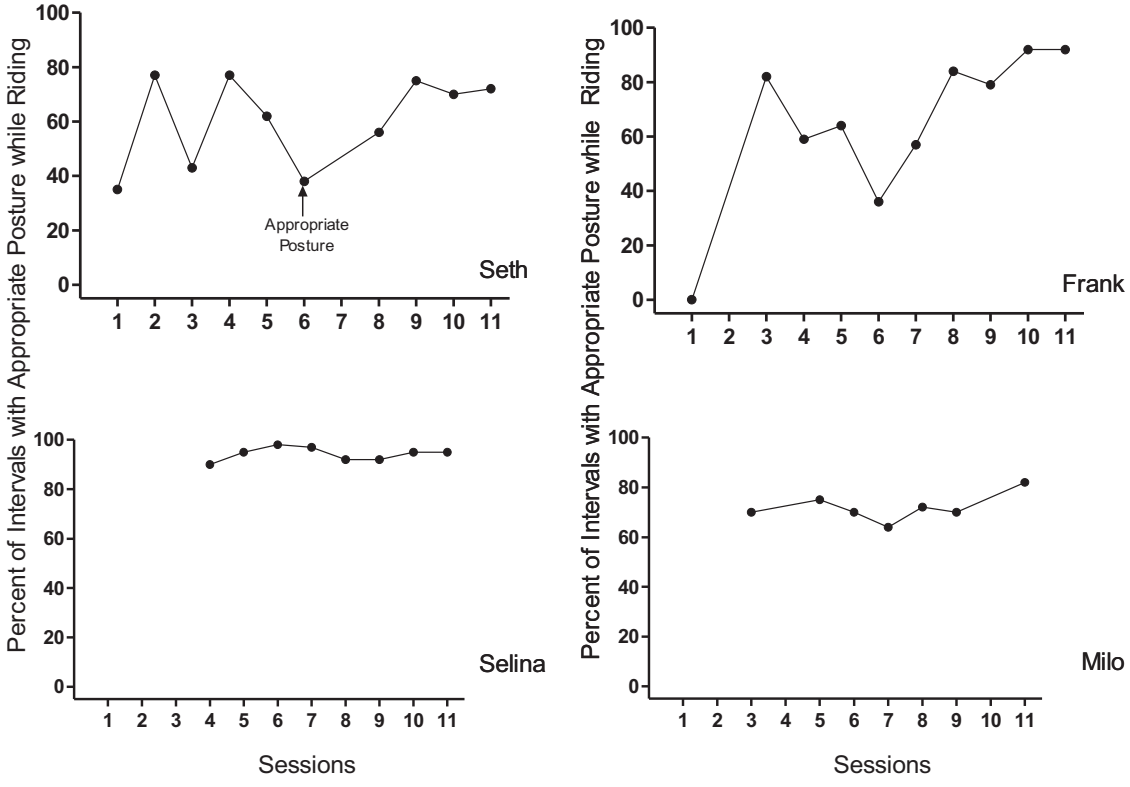

Fig. 9. Percentage of intervals with appropriate posture for the treatment group during THR.

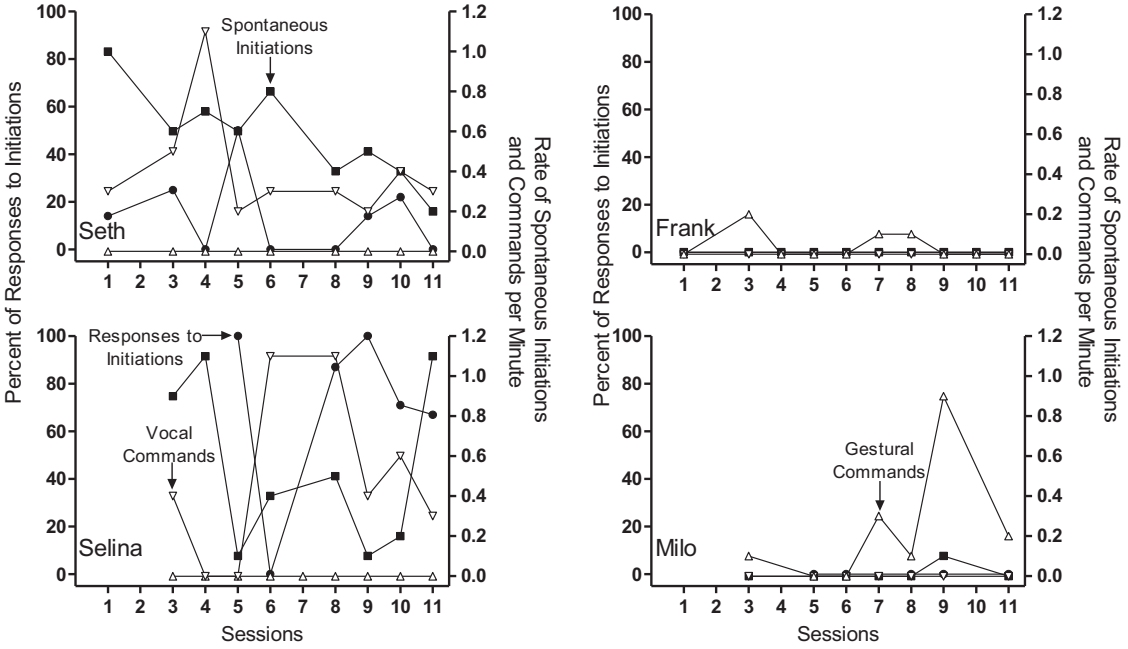

Fig. 10. Rate of spontaneous initiations and vocal and gestural commands given to the horse, and percentage of responses to others' initiations for the treatment group during THR. Closed squares indicate spontaneous initiations, open upside-down triangles indicate vocal commands, open triangles indicate gestural commands, and closed circles indicated responses to initiations.

2.2.4. Vocal commands

The rate of vocal commands used to direct the horse was captured via a voice recorder, in which a 10 min time sample of each THR session was scored for each participant. These data are depicted in Fig. 10. Seth's rate of vocal commands was variable with a one-session increase in the rate of vocal commands ($M = 0.4$; range, 0.2–1.1). Selina's rate was also variable and showed no clear trend ($M = 0.49$; range, 0–1.1). Neither Frank nor Milo used vocal commands during THR. These data suggest that the rate of vocal commands remained unchanged throughout THR.

2.2.5. Gestural commands

The rate of gestural commands was recorded in vivo during each THR session for 10 min per participant. These data are depicted in Fig. 10. Neither Seth nor Selina used gestural commands during therapy. Frank's rate of gestural commands was low and somewhat stable ($M = 0.04$; range, 0–0.2). Milo's rate of gestural commands was variable ($M = 0.23$; range, 0–0.9). Overall, these data indicate little changes in the rate of gestural commands during THR.

2.2.6. Compliance

The percentage of compliance was recorded prior to participants stepping in the arena (e.g., putting on the helmet, mounting the horse) with their horses as well as when they dismounted (e.g., dismounting the horse without resistance, removing their helmet). These data are depicted in Fig. 11. The percentage of compliance for Seth was stable initially and increased in variability thereafter ($M = 81.5\%$; range, 25–100%). Selina complied with all directions ($M = 100\%$) during THR. Frank's percentage showed an increasing trend with stability observed at 80% by the end of the study ($M = 72.5\%$; range, 25–100%). Milo had a high and stable percentage of compliance ($M = 94.29\%$; range, 80–100%). In sum, increases in compliance were observed for two of four participants during THR while no changes were demonstrated for the remaining participants (though compliance, on average, was high).

2.2.7. Problem behavior

The percentage of intervals in which participants engaged in problem behavior was recorded for 10 min in vivo (e.g., motor stereotypy) as well as from voice recorders (e.g., vocal stereotypy). These data are depicted in Fig. 11. Seth's percentage of problem behavior was low and stable ($M = 4.6\%$; range, 0–11%). Selina did not engage in problem behavior during THR. Frank's percentages were variable ($M = 48.82\%$; range, 3–85%). Milo's problem behavior was low and stable ($M = 5.4\%$; range, of 2–12%). In sum, three of four participants displayed low levels of problem behavior during THR.

2.3. CBCL and teacher rating form (TRF)

Parents and teachers completed rating forms regarding behavior problems and competencies their child/student displayed before the study started and after its completion. Data for the CBCL and TRF for internalizing problems, externalizing problems, and total problems for the treatment and control group are reported in Tables 3 and 4, respectively.

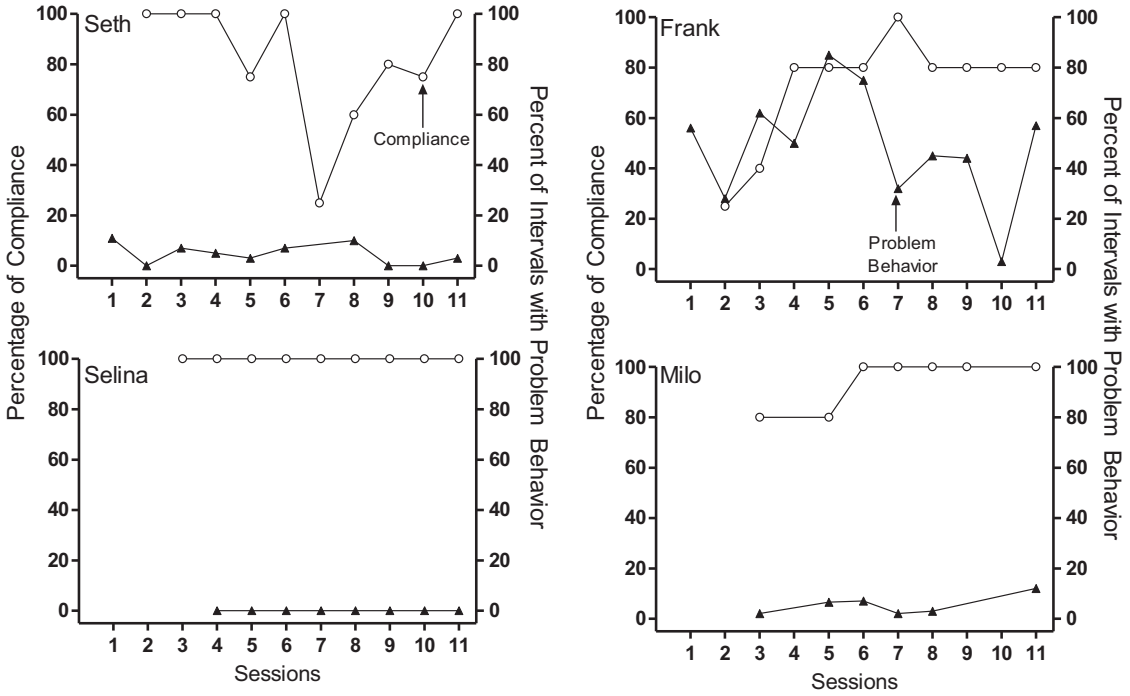

Fig. 11. Percentage of intervals with problem behavior and the percentage of compliance for the treatment group during THR. Closed triangles indicate problem behavior and open circles indicate compliance during THR sessions.

Two of the participants in the treatment group (Seth and Selina) did not show meaningful changes in their pre-post internalizing problems, externalizing problems, and total problems *T*-scores for the CBCL. Frank's *T*-scores increased over time and were within clinical ranges by the end of the study. Milo's *T*-scores for externalizing problems and total problems improved by the end of the study and were within normal ranges. All of the control group participants showed improved

Table 3
Treatment and control group pre- and post-study CBCL scores.

Group	Pre-study CBCL *T*-score	Behavior range	Post-study CBCL *T*-score	Behavior range
Treatment group				
Seth				
Internalizing problems	45	Normal	45	Normal
Externalizing problems	58	Normal	54	Normal
Total problems	55	Normal	53	Normal
Selina				
Internalizing problems	39	Normal	33	Normal
Externalizing problems	56	Normal	52	Normal
Total problems	51	Normal	47	Normal
Frank				
Internalizing problems	48	Normal	66	Clinical
Externalizing problems	60	Borderline	68	Clinical
Total problems	62	Borderline	73	Clinical
Milo				
Internalizing problems	52	Normal	50	Normal
Externalizing problems	63	Borderline	54	Normal
Total problems	64	Clinical	57	Normal
Control group				
Ivan				
Internalizing problems	62	Clinical	59	Normal
Externalizing problems	64	Clinical	59	Normal
Total problems	70	Clinical	64	Clinical
Denis				
Internalizing problems	50	Normal	48	Normal
Externalizing problems	53	Normal	63	Borderline
Total problems	54	Normal	66	Clinical
Edmund				
Internalizing problems	72	Clinical	60	Borderline
Externalizing problems	67	Clinical	50	Normal
Total problems	73	Clinical	53	Normal

Table 4
Treatment and control group pre- and post-study TRF score.

Group	Pre-study TRF *T*-score	Behavior range	Post-study TRF *T*-score	Behavior range
Treatment group				
Seth				
Internalizing problems	58	Normal	45	Normal
Externalizing problems	55	Normal	58	Normal
Total problems	61	Borderline	56	Normal
Selina				
Internalizing problems	57	Normal	64	Clinical
Externalizing problems	63	Borderline	66	Clinical
Total problems	63	Borderline	62	Borderline
Frank				
Internalizing problems	59	Normal	62	Borderline
Externalizing problems	69	Clinical	66	Clinical
Total problems	70	Clinical	67	Clinical
Milo				
Internalizing problems	50	Normal	50	Normal
Externalizing problems	42	Normal	54	Normal
Total problems	53	Normal	56	Normal
Control group				
Ivan				
Internalizing problems	52	Normal	52	Normal
Externalizing problems	54	Normal	56	Normal
Total problems	54	Normal	56	Normal
Denis				
Internalizing problems	53	Normal	59	Normal
Externalizing problems	71	Clinical	70	Clinical
Total problems	72	Clinical	73	Clinical
Edmund				
Internalizing problems	60	Borderline	64	Clinical
Externalizing problems	53	Normal	55	Normal
Total problems	52	Normal	56	Normal

T-scores that were within normal ranges (Ivan and Edmund) by the end of the study. The externalizing and total problems *T*-scores for Denis increased over time and were within borderline and clinical ranges, respectively, at the conclusion of the study. These findings suggest that THR did not produce systematic and meaningful improvement in the areas measured by the CBCL (Table 3).

One of the participants in the treatment group (Milo) did not show meaningful changes in his pre-post internalizing problems, externalizing problems, and total problems *T*-scores for the TRF. His scores were within the normal range before and after THR. The internalizing and externalizing problems *T*-scores remained in the normal range for Seth after THR; however, the total problems *T*-score improved. Two of Selina's *T*-scores increased to clinical ranges post-THR (internalizing and externalizing problems). Her total problems *T*-score remained in the borderline range. Frank's *T*-scores remained in the same clinical range (externalizing and total problems) or increased to the borderline range (internalizing problems) after THR. Two of the control group participants showed no change in their pre-post TRF *T*-scores for the internalizing, externalizing, or total problems scales (Ivan and Denis). Edmund's *T*-scores for externalizing and total problems remained in the normal range at the end of the study; however, his *T*-scores for internalizing problems increased to the clinical range. These findings suggest that THR did not produce systematic and meaningful improvement in the areas measured by the TRF (Table 4).

Table 5
Post-THR parent survey.

Question	*M*	SD
Therapeutic horseback riding helped to increase my child's use of language (sign language, speech, or use of a language device)	2.25	0.96
Therapeutic horseback riding helped decrease the frequency of problem behavior (elopement, aggression, pica, etc.) that my child displays	2.0	0.82
I think therapeutic horseback riding was a fun activity for my child to experience	4.0	0
My child's motivation increased	3.33	0.58
My child appeared to be excited to participate in therapeutic horseback riding	3.75	0.05
Therapeutic horseback riding helped to improve my child's level of independence	2.5	1.29
Therapeutic horseback riding was very beneficial to my child's overall skill and behavior acquisition	3.33	0.58
My child will continue to participate in therapeutic horseback riding	2.67	0.58

Note: 1 = strongly disagree, 4 = strongly agree.

2.4. Survey data

Parents completed a survey after the conclusion of THR to assess their opinions of the effects of THR for their child. Higher scores represent agreement regarding benefits of THR. Parents did not agree with statements that THR (1) helped to increase their child's use of language ($M = 2.25$; range, 1–3), (2) was an effective intervention to decrease problem behavior ($M = 2$; range, 1–3), or (3) improved their child's level of independence ($M = 2.5$; range, 1–4). Parents slightly agreed that THR improved their child's motivation ($M = 3.33$; range, 2.5–4) and was beneficial to their child's overall skill and behavior acquisition ($M = 3.33$; range, 2.5–4); however, when asked if their child would continue to participate in THR sessions, the average rating did not reflect this perceived benefit ($M = 2.7$; range, 2–3). Parents agreed that their child appeared excited to participate ($M = 3.75$; range, 3–4) and that THR was a fun activity for their child to experience ($M = 4$). Open-ended comments indicated that some parents would look into other opportunities for their child to continue THR. Others noted that their child loved going each week and that they noticed an increase in language expression. These data are depicted in Table 5.

3. Discussion

The purpose of the present study was to use a single-case experimental design to evaluate the effects of THR on numerous behaviors of children with ASD. In addition, this study included a waitlist control group for comparison purposes. The results suggest that THR did not produce clinically significant effects on participant affect, off-task behavior, problem behavior, compliance, or language (i.e., spontaneous initiations and responses to initiations) from baseline to treatment during center-based activities and home observations. The findings from the time-series analysis suggest that THR is not an effective intervention to improve performance on these dependent variables. Data for participants who received THR are similar to participants assigned to the waitlist control group, in that data showed similar variability (or stability) across many, if not most, of the dependent variables for these settings. Improvements were noted for posture during THR sessions, but a lack of experimental control during THR prevents demonstration of a functional relation. Scores on the CBCL showed reductions in problem behavior at the end of the study for both treatment and control group participants indicating that THR was not responsible for observed reductions in the treatment group. Interestingly, the TRF scores differ from the CBCL scores and showed little change before and after THR. Post-THR parent surveys appear to support the main findings. Although parents indicated appreciation for the opportunity for their child to participate in THR, it was not perceived as an effective therapy for addressing problem behaviors or language deficits. Interestingly, anecdotal verbal reports from some parents indicated changes in their child's language both at home and at school; however, these changes were not captured during data collection over time and across settings. Parents appear to indicate that THR may be best conceptualized as a leisure activity, rather than a treatment option for symptoms of ASD. In summary, these findings indicate that THR did not provide therapeutic benefit to the participating children with autism.

Past research has evaluated the effects of THR on various dependent variables using pre-post assessments, surveys, waitlist control groups, and in one case, a reversal design. The collective results of these studies suggest that THR improved social skills, fine motor perception (Bass et al., 2009; Nelson et al., 2011), motor skills, and self-regulation (Gabriels et al., 2012) across participants. While these past results suggest promising effects of THR, the results of the current study not only fail to replicate the magnitude of effects across dependent variables, but also fail to demonstrate any meaningful improvements across participants. Perhaps this failure to replicate may be due to the way in which data were collected. While past studies relied on self-report forms and poorly defined constructs, the current study used direct observation of operationally defined dependent variables as well as repeated measurement. Although the current study relied on some self-report forms, they were used as a *supplement to*, not in lieu of, direct observation techniques, which are the hallmark of behavior analytic research. It may also be the case that participants in this study differed in a meaningful way from participants in previously published research (e.g., skill and functioning level, experiences outside of THR sessions). Additionally, the number of sessions of THR offered to the participants differs across studies and may be responsible for the differences in treatment effects. It is important to note, however, that participants were exposed to 1 h sessions for 9 weeks, which is still within the range of exposure of other studies (2.5 h to 6 months). The content of each lesson may also be a contributing factor for the failure to replicate. The various procedures used and opportunities to practice different skills may influence skill acquisition and/or behavior change; however, there is no research thus far to suggest that a specific set of lesson plans for THR is superior to, or more effective, than another.

The current study adds to the existing body of literature by using a single-case experimental design to evaluate the effectiveness of THR on target behaviors that are representative of the outcomes evaluated in prior studies. Moreover, this study evaluated an already established program that was not created or implemented by the researchers, which eliminates the skeptical bias that may be present when researchers evaluate alternative treatments that do not have strong evidence to support their use. This study contradicts previous research and failed to show benefits across the dependent variables. As the number of ASD treatments continues to increase, more experimental research is needed to evaluate these treatment options. This will not only help parents and practitioners select the most appropriate intervention, but will hopefully preserve valuable instructional or treatment time. In addition, this study will help to inform EBP, which provides some guidance about which interventions are most appropriate for a particular clinical issue. Identifying interventions that *are not* effective, or even harmful, is as important as documenting interventions that *are* effective (Romanczyk et al., 2008).

Although this study has strengths and addresses a gap in the literature, it also has several limitations. First, the research team relied on video cameras to capture center activities that were not scored in vivo. For a few sessions, the cameras lost battery power and did not record sessions for some participants. These sets of data were not recovered and, thus, were not included in the analysis of the effects of THR. However, because dependent variables were measured two to four times weekly during center activities, this data loss was minimal and may have only marginally impacted data analysis.

Next, dependent variables were measured for one participant at a time during therapy. Those participants most likely to dismount before the end of the session due to problem behavior were observed first. Consequently, Selina, who was the most likely to remain on the horse during the entire session, was not always observed for a full 10 min (despite riding for the whole session). This was the case even if the lesson began late because lessons had to end at a specific time due to volunteer availability. Since some dependent variables were recorded in vivo during these sessions, those data were not captured; however, data for a vast majority of Selina's therapy sessions were recorded. Given the lack of change in behavior throughout THR sessions, it is unlikely that this loss of data greatly impacted data analysis. Furthermore, Seth and Frank were dismounted before the conclusion of a lesson on two separate occasions due to unsafe behavior and/or non-assent, which may have impacted their horsemanship skill acquisition. To mitigate this limitation, both participants were given the opportunity to ride a horse during two additional sessions, which equated to a total of nine full sessions of therapy. Additionally, because of missed sessions, time, and resource constraints during the study, Milo did not complete nine full sessions of THR. The therapy facility and research team were unable to offer additional sessions. This may have impacted skill acquisition during therapy and in other settings; however, given the data from other participants, it is less unlikely that these absences greatly impacted Milo's behavior.

THR was offered for 9 weeks, which is a shorter amount of exposure time than other studies. Perhaps this is a limiting factor which may have impacted skill acquisition. Future studies may be able to capitalize on the single-case design methodology by measuring dependent variables for an extended period of time, such as 12–16 weeks. This extended time would allow for an examination of trends in behavior over a longer period of time as well as reduce the likelihood that behavior changes occur due to extraneous variables, while also evaluating the importance of "dose" with regard to THR exposure. Another area for future research includes evaluating the use of THR as a reinforcer for performance during teaching sessions. For example, after completing a token board during discrete trial instruction, a student could be allowed to participate in a session of THR. Instructors or teachers could capitalize on the Premack principle, in which performance during work sessions may improve because it is followed by a highly preferred activity. THR could be used as a reinforcer within a treatment package, but should not be used as the primary treatment option to change behavior (given the results of the present study).

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Child behavior checklist for ages 6 to 18*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Atkins, R. S., Angkustsiri, K., & Hansen, R. L. (2010). Complementary and alternative medicine in autism: Evidence-based approach to negotiating safe and efficacious interventions with families. *Journal of the American Society for Experimental NeuroTherapeutics, 7*, 307–319.

Association for Science in Autism Treatment. (n.d.). *Summaries of scientific research on interventions on autism*. Retrieved April 1, 2012, from http://www.asatonline.org/treatment/treatments_desc.

Autism Speaks. (n.d.a). *Congress briefed on rising prevalence, cost of autism*. Retrieved June 18, 2012, from http://www.autismspeaks.org/blog/2012/05/02/congress-briefed-rising-prevalence-cost-autism.

Autism Speaks. (n.d.b). *Facts about autism*. Retrieved June 18, 2012, from http://www.autismspeaks.org/what-autism/facts-about-autism.

Autism Speaks. (n.d.c). *What is autism? What is autism spectrum disorder*? Retrieved June 13, 2012, from http://www.autismspeaks.org/what-autism.

Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders, 39*, 1261–1267 http://dx.doi.org/10.1007/s10803-009-0734-3.

Bertoti, D. B. (1988). Effect of therapeutic horseback riding on posture in children with cerebral palsy. *Physical Therapy, 68*, 1505–1511.

Bracher, M. (2000). Therapeutic horse riding: What has this to do with occupational therapists? *British Journal of Occupational Therapy, 63*, 277–282.

Bruininks, R. H., & Bruininks, B. D. (2005). *Bruininks-Oseretsky test of motor proficiency* (2nd ed.). Minneapolis, MN: Pearson Assessment.

Centers for Disease Control and Prevention. (2011). *Facts about autism* Retrieved September 1, 2011, from http://www.cdc.gov/ncbddd/autism/facts.html.

DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40*, 447–461 http://dx.doi.org/10.1901/jaba.2007.40-447.

DiGennaro Reed, F. D., & Reed, D. D. (2008). Towards an understanding of evidence-based practice. *Journal of Early and Intensive Behavior Intervention, 5*, 20–29.

Drnach, M., O'Brien, P. A., & Kreger, A. (2010). The effects of a 5-week therapeutic horseback riding program on gross motor function in a child with cerebral palsy: A case study. *The Journal of Alternative and Complementary Medicine, 16*, 1003–1006 http://dx.doi.org/10.1089/acm.2010.0043.

Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., et al. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*, 578–588 http://dx.doi.org/10.1016/j.rasd.2011.09.007.

Green, C. W., & Reid, D. H. (1999). A behavioral approach to identifying sources of happiness and unhappiness among individuals with profound multiple disabilities. *Behavior Modification, 23*, 280–293 http://dx.doi.org/10.1177/0145445599232006.

Kay, S., & Vyse, S. (2008). Helping parents separate the wheat from the chaff: Putting autism treatments to the test. In J. W. Jacobson, R. M. Foxx, & J. A. Mulick (Eds.), *Controversial therapies for developmental disabilities: Fad, fashion, and science in professional practice* (pp. 129–138). Hove, East Sussex: Psychology Press.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*, 146–159 http://dx.doi.org/10.1037/0003-066X.63.3.146.

Kern, J. K., Fletcher, C. R., Mehta, J. A., Grannemann, B. D., Knox, K. R., Richardson, T. A., et al. (2011). Prospective trial of equine-assisted activities in autism spectrum disorder. *Alternative Therapies, 17*(3), 14–20.

Matson, J. L., Sevin, J. A., Fridley, D., & Love, S. R. (1990). Increasing spontaneous language in three autistic children. *Journal of Applied Behavior Analysis, 23*, 227–233 http://dx.doi.org/10.1901/jaba.1990.23-227.

National Standards Project. (2009). *Evidence-based practice and autism in the schools: A guide to providing appropriate interventions to students with autism spectrum disorders*. Randolph, MA: National Autism Center.

Nelson, K., Axtell, J., Derby, K. M., Moug, R., Berrera, S., & McLaughlin, T. F. (2011). A preliminary analysis of therapeutic horseback riding. *International Journal of Social Sciences and Education, 1*(4), 644–656.

Nickel, R. E., & King Gerlach, E. (2001). Use of complementary and alternative therapies by the families of children with chronic phases and disabilities. *Infants and Young Children, 14*, 67–78.

Professional Association of Therapeutic Horsemanship International. (n.d.). *Path International*. Retrieved June 1, 2012, from http://www.pathintl.org/.

Romanczyk, R. G., Gillis, J. M., White, S., & DiGennaro, F. D. (2008). Comprehensive treatment packages for ASD; perceived vs proved effectiveness. In J. Matson (Ed.), *Clinical assessment and intervention for autism spectrum disorders*. Burlington, MA: Elsevier Inc.

Shafer, M. S., Egel, A. L., & Neef, N. A. (1984). Training mildly handicapped peers to facilitate changes in the social interaction skills of autistic children. *Journal of Applied Behavior Analysis, 17*, 461–476 http://dx.doi.org/10.1901/jaba.1984.17-461.

Schopler, E., Reichler, R. J., & Renner, B. R. (1994). *The childhood autism rating scale*. Los Angeles, CA: California Western Psychological Services.

Snider, L., Korner-Bitensky, N., Kammann, C., Warner, S., & Saleh, M. (2007). Horseback riding as therapy for children with cerebral palsy: Is there evidence of its effectiveness? *Physical & Occupational Therapy in Pediatrics, 27*, 5–23 http://dx.doi.org/10.1300/J006v27n02_02.

Umbarger, G. T. (2007). State of the evidence regarding complimentary and alternative medical treatments for autism spectrum disorders. *Education and Training in Developmental Disabilities, 42*, 437–447.

Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39*, 369–373 http://dx.doi.org/10.1901/jaba.2006.144-05.

Wong, H. H. L., & Smith, R. G. (2006). Patterns of complementary and alternative medical therapy use in children diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disabilities, 36*, 901–909 http://dx.doi.org/10.1007/s10803-006-0131-0.