WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2017

Evaluating Treatment Integrity Failures during Timeout from Play

Apral P. Foreman

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation

Foreman, Apral P., "Evaluating Treatment Integrity Failures during Timeout from Play" (2017). *Graduate Theses, Dissertations, and Problem Reports*. 5612.
https://researchrepository.wvu.edu/etd/5612

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Evaluating Treatment Integrity Failures during Timeout from Play

Apral P. Foreman

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Psychology

Claire St. Peter, Ph.D., BCBA-D, Chair
Karen G. Anderson, Ph.D.
Regina Carroll, Ph.D., BCBA-D
Amy Herschell, Ph.D.
Alexandra Hollo, Ph.D., BCBA-D

Department of Psychology

Morgantown, West Virginia
2017

Keywords: Timeout; Treatment Integrity; Children; Problem Behavior
Copyright 2017 Apral P. Foreman

ABSTRACT

Evaluating Treatment Integrity Failures during Timeout from Play

Apral P. Foreman

Timeout is an effective behavior-reduction strategy with considerable generality. There are several recommendations about how to implement timeout. However, little research has investigated how timeout is implemented under natural conditions, or how timeout-implementation errors impact its effectiveness. Thus, our study attempted to address two aims. The first aim was to observe how teachers implemented timeout with their students. To address this aim (Experiment 1), we completed naturalistic observations of teachers who were implementing timeout during play. We collected data on how frequently the teachers implemented timeout following problem behavior (omission integrity), and how frequently teachers implemented timeout following responses not targeted for timeout (commission integrity). Experiment 1 data showed that teachers rarely implemented timeout; when the teachers did implement timeout, they often did not follow the timeout parameters specified initially. The second aim was to evaluate what effects inconsistent timeout implementation has on student behavior. To address this aim (Experiment 2), we evaluated the effects of inconsistent timeout on rates of problem behavior. Specifically, we used a reversal design to compare implementation of timeout at 0%, 100%, and reduced integrity (e.g., 11% integrity). The specific level of integrity implemented during the reduced-integrity phases was based upon the omission integrity observed for each student during Experiment 1. The data from Experiment 2 showed that timeout implemented with high integrity decreased problem behavior for two of four students. For one student, reduced-integrity implementation also decreased problem behavior. Due to low rates of problem behavior and participant attrition, three of four participants did not complete Experiment 2. Although our conclusions are limited, these data allowed us to make clinical recommendations about how the teachers should implement timeout with their students. Our results also have important implications for the application of timeout in schools and for future research.

Table of Contents

Introduction 1

Experiment 1 11
- **Method** 11
- **Results** 21
- **Discussion** 25

Experiment 2 27
- **Method** 27
- **Results** 36
- **Discussion** 39

General Discussion 41

References 47

Tables 55

Figures 62

Appendix A 66

Appendix B 68

Appendix C 75

Appendix D 76

Evaluating Treatment Integrity Failures during Timeout from Play

Timeout involves loss of access to reinforcers contingent upon undesirable behavior, such as aggression or inappropriate vocalizations (e.g., Cuenin & Harris, 1986; Hobbs & Forehand, 1977; Warzak, Floress, Kellen, Kazmerski, & Chopko, 2012). In general, timeout should result in a shift from a reinforcing context containing reinforcing activities and attention, to a less reinforcing context, in which access to the previously available reinforcers is absent for a brief period (Brantner & Doherty, 1983). Timeout, when used correctly and appropriately, should reduce undesirable behavior. Given this broad definition, timeout can be implemented in several different ways (Brantner & Doherty, 1983; Harris, 1985; Ritz, Noltemeyer, Davis, & Green, 2014).

Timeout procedures are often categorized as isolation, exclusion, or nonexclusion timeout (Brantner & Doherty, 1983; Harris, 1985). Isolation timeout involves completely removing the individual, usually to another room, from the reinforcing context. Exclusion timeout typically involves removing the individual to another location in the same area (i.e., not to another room) as the reinforcing context. For example, a child may be told to sit in a chair facing a corner of the classroom (Brantner & Doherty, 1983). Nonexclusion timeout involves the individual remaining in the reinforcing context, but he or she is not allowed to access reinforcers. Nonexclusion timeout has also been referred to as contingent observation (e.g., Porterfield, Herbert-Jackson, & Risley, 1976) or a “Sit-and-Watch” procedure (e.g., White & Bailey, 1990). Regardless of categorization, timeout procedures produce robust effects with considerable generality (e.g., Barton, Guess, Garcia, & Baer, 1970; Crepsi, 1988; Donaldson & Vollmer, 2011; Donaldson, Vollmer, Yakich, & Van Camp, 2013; Fabiano et al., 2004; Foxx & Shapiro,

1978; McKeegan, Estill, & Campbell, 1984; Salend & Gordon, 1987; White, Nielsen, & Johnson, 1972).

Timeout is an effective and frequently used behavior-management strategy for preschool-aged children. For example, Porterfield, Herbert-Jackson, and Risley (1976) compared the effects of timeout and a redirection strategy on disruptive behavior of toddlers during play. They found that timeout was more effective in suppressing disruptive behavior than the redirection strategy. As another example, Donaldson and Vollmer (2011) demonstrated that 4-min timeouts reduced the challenging behavior of 3-and 4-year-old children.

Timeout is also effective for older students. For example, Spitalnik and Drabman (1976) evaluated timeout with a 12- and 14-year-old student in a residential facility classroom by comparing a baseline condition to a timeout condition. They found a 50% decrease in vocalizations during the timeout condition relative to baseline. McKeegan, Estill, and Campbell (1984) reduced stereotypy of a 28-year-old man diagnosed with Autism by implementing 2-min timeouts.

The effectiveness of timeout has been demonstrated with individuals diagnosed with various disabilities. Olmi, Sevier, and Nastasi (1997) suppressed the challenging behavior of a child with a learning disability and a child diagnosed with cerebral palsy. Both Fabiano et al. (2004) and Kapalka and Bryk (2007) demonstrated the effectiveness of timeout with boys with Attention-Deficit/Hyperactivity Disorder (ADHD) diagnoses. Foxx and Shapiro (1978) used 3-min timeouts to reduce the disruptive behavior of boys with intellectual disabilities. Timeout's effectiveness has also been demonstrated in a variety of settings. A few include psychiatric hospitals (e.g., Bostow & Bailey, 1969; Burchard & Barrera, 1972; McGuffin, 1991), classrooms (e.g., Huguenin, 1981; Marlow, Tingstrom, Olmi, & Edwards, 1997; Wilson, Robertson,

Herlong, & Haynes, 1979), and outpatient or clinic settings (e.g., Everette, Olmi, Edwards, Tingstrom, Sterling-Turner, & Christ, 2007; Warzak & Floress, 2009).

Although the effectiveness of timeout has been replicated across numerous studies, its robust effects depend on several parameters (Brantner & Doherty, 1983; Harris, 1985; MacDonough & Forehand, 1973). For timeout to be successful, certain aspects of the procedure should be considered carefully. Several research studies have been conducted to determine the importance of certain timeout parameters, and how they should be implemented in practice.

Timeout parameters commonly researched include release criteria (e.g., Bean & Roberts, 1981; Donaldson & Vollmer, 2011; Mace, Page, Ivancic, & O'Brien, 1986), duration (e.g., Ferster & Appel, 1961; Hobbs, Forehand, & Murray, 1978; Kaufman & Baron, 1968, Experiment 2; White et al., 1972; Zimmerman & Baydan, 1963, Experiment 1), and schedule (e.g., Clark, Rowbury, Baer, & Baer, 1973; Zimmerman & Baydan, 1963, Experiment 2). General recommendations for timeout implementation based on these parameters have been developed from the literature. These recommendations include the use of brief timeout durations (i.e., 5 min or less) and timeout implementation following each instance of problem behavior (i.e., continuous schedule; Brantner & Doherty, 1983; Harris, 1985; MacDonough & Forehand, 1973) despite some evidence that intermittent timeout can be effective.

Several studies recommend implementing timeout following every target response (i.e., on a continuous schedule; Brantner & Doherty, 1983; Pendergrass, 1971; Zimmerman & Baydan, 1963). For example, Pendergrass (1971) found that brief timeouts implemented on a continuous schedule effectively suppressed the problem behavior of a 5-year-old girl with brain damage; intermittent timeouts were ineffective even when implemented for a longer duration of time.

In contrast, other research suggests that some intermittent timeout schedules may be effective at suppressing problem behavior to low rates (e.g., Barton, Brulle, & Repp, 1987; Clark et al., 1973, Experiment 2; Donaldson & Vollmer, 2012; Jackson & Calhoun, 1977). Donaldson and Vollmer (2012) maintained low rates of problem behavior while thinning the timeout schedule from either a fixed-ratio (FR) 1 schedule or an FR 2 schedule to a variable-ratio (VR) 2 schedule for two students; for one student, an FR 1 schedule was thinned to a VR 5. Clark, Rowbury, Baer, and Baer (1973) evaluated effects of several intermittent timeout schedules on the disruptive behavior of an 8-year-old girl. When timeout occurred frequently (e.g., timeout followed each disruptive behavior), less disruptive behavior occurred than when timeout was less frequent (e.g., timeout followed approximately eight responses). However, the relation between the timeout frequency and rate of disruptive behavior was nonlinear. Disruptive behavior occurred at similar rates when timeout followed either each instance of behavior (FR 1) or approximately four instances (VR 4). When timeout followed approximately eight instances (VR 8), disruptive behavior occurred twice as much than when timeout was implemented on a VR 4 schedule; when timeout was not implemented at all (baseline), disruptive behavior occurred five times as much than when timeout was implemented on a VR 4. Donaldson and Vollmer and Clark et al.'s findings suggest that some intermittent timeout schedules may suppress problem behavior to low rates.

Jackson and Calhoun (1977) compared phases during which timeout was implemented on a VR 8 or VR 4 to phases in which no timeouts were implemented (baselines) with a 10-year-old boy diagnosed with an intellectual disability. They found intermittent timeouts to be effective in suppressing disruptive behavior. Calhoun and Matherne (1975) compared the effects of timeout on the aggressive behavior of a 7-year-old girl diagnosed with an intellectual disability. They

evaluated timeout implemented continuously (FR 1), on an FR 2, and on an FR 5. Both the FR 1 and FR 2 suppressed aggression relative to baseline phases; timeout implemented on an FR 5 was ineffective (Calhoun & Matherne, 1975). Similar to Clark et al. (1973), both Jackson and Calhoun and Calhoun and Matherne found that the richer of the two intermittent schedules (i.e., VR 4, FR 2) was more effective in decreasing aggression than the leaner of the two intermittent schedules (i.e., VR 8, FR 5).

The effectiveness of intermittent timeout schedules in decreasing target behavior has also been demonstrated in the laboratory. Zimmerman and Baydan (1963, Experiment 2) had two college students complete a matching-to-sample task during 2-hr sessions a few times per week. During the task, a sample stimulus was presented (e.g., a square). Pressing a key resulted in the removal of the sample stimulus (i.e., square) and the simultaneous presentation of two stimuli (e.g., a square and a circle). Sometimes these stimuli would be superimposed onto red stimuli. When the stimuli were red, the students were expected to select the stimulus that matched the sample stimulus; when the stimuli were not red, the students were expected to select the non-matching stimulus. Correct responses (i.e., selecting the correct stimulus given the sample and color of the comparison stimuli) resulted in intermittent, conditioned reinforcers (click sound, flash of a light, and increase in tallies on a counter). Incorrect responses resulted in a timeout (i.e., the apparatus turned off for a fixed duration of time – timeout duration ranged from 2 s to 120 s across conditions). Students were exposed to timeouts implemented on F1, FR 2, FR 5, FR 10, and FR 50 schedules. At the end of each 2-hr session, tallies were exchanged at a rate of $0.20 per tally.

Zimmerman and Baydan (1963) found that, holding timeout duration constant, incorrect responding increased as timeout became more intermittent. Leaner schedules (more infrequent

timeout implementation) resulted in less suppression of incorrect responses. In a similar preparation with pigeons, Zimmerman and Ferster (1963) demonstrated a similar relation between timeout frequency and suppression of incorrect responding on a matching-to-sample task. Timeout research in both the laboratory and applied settings suggests that intermittent timeout can be effective in decreasing behavior targeted for reduction, at least with relatively rich timeout schedules (e.g., FR 2, VR 3, etc.).

Despite effectiveness, intermittent timeout implementation could be considered inaccurate implementation. For example, if timeout is programmed on a continuous schedule, but implemented on a VR 5, implementation could be characterized as inconsistent. The extent to which procedures are accurately or consistently implemented as described is sometimes referred to as *treatment integrity* (Peterson, Homer, & Wonderlich, 1982). The integrity with which an intervention is implemented can impact the efficacy of the intervention (e.g., Arkoosh, Derby, Wacker, Berg, McLaughlin, & Barretto, 2007; Fryling, Wallace, & Yassine, 2012).

Treatment integrity evaluations are important because inaccurate implementation of an intervention can interfere with the assessment of the intervention's effectiveness. When treatment integrity is low, it may be difficult to determine whether the intervention is failing because the strategy is ineffective or if it is just not being implemented correctly. For example, you must first pull the pin on a fire extinguisher for it to work properly. If you forget to pull the pin, and try to use the extinguisher, it will not work. You might consider the extinguisher faulty, but in fact, the extinguisher is in perfect working order; it was just not used properly.

In the existing literature, timeout is often implemented with high levels of treatment integrity or consistency (i.e., on average greater than 90% integrity; Alberto, Heflin, & Andrews, 2002; Fee, Matson, & Manikam, 1990; Mace et al., 1986; Marlow et al., 1997). In other words,

timeout is programmed to be implemented on a certain schedule and the implementation of the programmed schedule is fairly accurate. Thus, it is unclear how reduced treatment integrity (or inconsistent implementation) influences the efficacy of timeout.

To evaluate effects of intermittent (or reduced integrity) implementation, the consistency of implementation can first be observed naturalistically through descriptive observations. For example, Carroll, Kodak, and Fisher (2013, Experiment 1) observed the delivery of academic instruction to children diagnosed with Autism Spectrum Disorder (ASD). They collected data on the consistency of various aspects of instruction delivery (e.g., giving a clear instruction, praising correct responses, etc.). From these naturalistic observations, Carroll et al. (2013) found that teachers were the most inconsistent in the delivery of tangibles following correct responses, providing prompts to evoke correct responses, and presenting the instruction only once. These results were then used to inform their second and third studies that systematically manipulated the integrity (or consistency) of these common errors.

Numerous studies have also experimentally manipulated treatment integrity to evaluate effects of reduced integrity on treatment outcomes (e.g., Carroll, Kodak, & Fisher, 2013, Experiment 2; DiGennaro Reed, Reed, Baez, & Maguire, 2011; St. Peter Pipkin, Vollmer, & Sloman, 2010; Wilder, Atwell, & Wine, 2006). In these studies, integrity failures are often categorized as omission or commission errors. Omission errors are said to occur when the therapist does not implement a component of the intervention. For example, St. Peter Pipkin, Vollmer, and Sloman (2010, Experiment 2) sometimes failed to provide scheduled attention following on-task behavior. In other words, scheduled attention was omitted from the intervention. Commission errors are said to occur when the therapist either implements a component of the procedure at the incorrect time or adds a component that is not part of the

intervention. For example, DiGennaro Reed, Reed, Baez, and Maguire (2011) sometimes provided tokens and praise following incorrect responses during a discrete-trial training (DTT) program.

Although several studies have evaluated how integrity failures impact differential-reinforcement or skill-acquisition procedures, only a few studies have evaluated how reduced treatment integrity influences the efficacy of timeout (e.g., Northup, Fisher, Kahang, Harrell, & Kurtz, 1997; Rhymer, Evans-Hampton, McCurdy, & Watson, 2002; Taylor & Miller, 1997). Taylor and Miller (1997, Experiment 1) demonstrated that higher treatment integrity with timeout procedures was related to greater suppression in problem behavior for two students with developmental disabilities. The experimenters measured how well classroom staff implemented timeout, and how often students engaged in problem behavior before and after staff training. Prior to training (baseline), staff implemented timeout with approximately 67% integrity, and the students engaged in more problem behavior. After training the staff to implement with approximately 98% integrity, the students engaged in less problem behavior. However, the experimenters did not directly manipulate treatment integrity levels.

Northup, Fisher, Kahang, Harrell, and Kurtz (1997) manipulated levels of treatment integrity during a multicomponent intervention consisting of differential reinforcement of alternative behavior (DRA) and timeout. They evaluated the multicomponent intervention (DRA plus timeout) at 100%, 50%, and 25% integrity. They also evaluated only the timeout component at 50% and 25% integrity while implementing the DRA component with 100% integrity. Finally, for one participant, they evaluated only the timeout component at 25% integrity plus extinction for appropriate behavior. Northup et al. found that, for two participants, DRA plus timeout remained effective until integrity for the multicomponent intervention (both

DRA and timeout) was at 25%. For the third participant, treatment effects maintained until the integrity of the timeout-only component was at 25% (while DRA was implemented with 100% integrity). When only timeout was implemented with 25% integrity and appropriate behavior was on extinction, inappropriate behavior increased and appropriate behavior decreased. In other words, the treatment effects were reduced when timeout was implemented with low integrity and appropriate behavior did not result in reinforcement. Given that the researchers only evaluated the timeout component alone for one participant, replication of the efficacy of timeout without a DRA treatment component is warranted.

Rhymer, Evans-Hampton, McCurdy, and Watson (2002) implemented a package intervention that included timeout at 25%, 50%, 75%, and 100% integrity with an 18-month-old girl who engaged in aggression at preschool. The timeout procedure involved four components: a reprimand following aggression (e.g., "No hitting."), a 15-30 s timeout following aggression, reinforcement of functionally equivalent appropriate behavior (e.g., obtaining adult attention by gently touching arm), and reinforcing all other appropriate behavior (e.g., praise for playing appropriately). To manipulate integrity, some number of components was randomly selected for implementation. For example, during 50% integrity only two of the four intervention components were implemented.

Rhymer et al. (2002) found that 75% and 100% integrity resulted in the best suppression of aggression. The 50% condition reduced aggression relative to baseline, but not as effectively as the 75% and 100% conditions. The 25% condition also reduced aggression relative to baseline, but there was much more session-to-session variability during the 25% condition. This study demonstrates that the treatment integrity of an intervention package including timeout can influence its effectiveness. However, the research design used was an ABCDEF (Baseline, 50%,

25%, 75%, 100%, Follow-Up) with one participant; none of the conditions were replicated, greatly limiting the study's internal validity. Thus, further timeout-integrity research is warranted.

The evaluation of omission errors during timeout could be considered an evaluation of intermittent timeout. Both involve timeout implemented on a non-continuous schedule. As mentioned previously, there is evidence to suggest that intermittent timeout can be effective. However, less is known about the effects of timeout implemented with reduced integrity. Although Taylor and Miller (1997, Experiment 1), Northup et al. (1997), and Rhymer et al. (2002) evaluated effects of reduced treatment integrity on timeout procedures, neither conduct both descriptive and experimental studies of reduced integrity.

Experimental studies of intermittent timeout have conflicting implications for timeout recommendations. For example, several studies recommend implementing timeout following every target response (e.g., Brantner & Doherty, 1983; Calhoun & Matherne, 1975; Pendergrass, 1971; Zimmerman & Baydan, 1963). However, other studies suggest that intermittent timeout may be effective in suppressing problem behavior to low rates (e.g., Clark et al., 1973, Experiment 2; Donaldson & Vollmer, 2012; Jackson & Calhoun 1977). Thus, descriptive timeout studies are needed to determine which of those recommendations are being followed in real-world applications of timeout and to further guide experimental timeout research.

Therefore, the current research had two specific aims. The first aim was to evaluate how treatment integrity failures occur naturally in an applied context. In Experiment 1, we evaluated naturally occurring omission and commission errors during play situations in a preschool and an elementary school. The second aim was to systematically investigate how treatment integrity failures impact the effectiveness of timeout. In Experiment 2, we manipulated the frequency of

omission errors during timeout procedures. We chose to manipulate only omission errors because they were more common in Experiment 1 than commission errors.

Experiment 1

Method

The purpose of Experiment 1 was to identify how frequently omission and commission errors occur during naturalistic implementation of timeout. We used results from Experiment 1 to determine omission-error parameters to manipulate during Experiment 2.

Recruitment and Interviews. We distributed information about the study to teachers or the director of the school. The teacher or director contacted the researcher to notify her of potential students and teachers that would be eligible for participation. A student and teacher were eligible for participation if the student engaged in problem behavior during play situations, and the teacher was implementing timeout during those play situations. No student and teacher recruited were deemed ineligible to participate.

After verifying that the student and teacher were eligible for participation, we obtained informed consent from the student's legal guardian and the teacher, and we obtained assent for students over 7 years old. Then, the researcher conducted structured, open-ended interviews with each teacher to determine what problem behavior the student engaged in during play and how the teacher was implementing timeout during play. The researcher asked teachers to describe behavior that resulted in timeout, as well as any other forms of problem behavior that the student emitted. The interview used is in Appendix A.

If the procedure was formally written in the student's Behavior Intervention Plan (BIP), the researcher also referred to the BIP when drafting definitions. The researcher brought any discrepancies between the BIP and the teacher's responses during the interview to the teacher's

attention. If the teacher said the BIP needed to be updated, the researcher waited until the BIP was updated before beginning data collection to ensure that the researcher and teacher were using the same definitions to record problem behavior.

Participants. Five students, aged 5 to 10 years old, who engaged in problem behavior during play situations and whose teachers were currently using timeout participated in Experiment 1. Willis was a 10-year-old boy who had diagnoses of ADHD, Reactive Attachment Disorder (RAD), and Seizure Disorder. Kyle was his teacher and was a long-term substitute teacher at Willis's school. He had previously worked with high-school students for 15 years. At the time of the interview with Kyle, timeout had been in place for Willis for 2 months.

Sonny was a 6-year-old boy who had diagnoses of Autism, Seizure Disorder, and ADHD. His teacher, Jill, had been teaching for 16 years, and was a Board Certified Behavior Analyst (BCBA). At the time of the interview with Jill, timeout had been in place for Sonny for 2 months.

Keith was a 9-year-old boy who had diagnoses of Traumatic Brain Injury (TBI), ADHD, and Bipolar Disorder. Keith was in a co-teach classroom; his teachers were Kelly and Cathy. Kelly had been teaching for 9 years and was a BCBA. Cathy had been teaching for 5 years. At the time of the interview with Kelly and Cathy, timeout had been in place for Keith for 1 month.

Ian was a 6-year-old boy who had diagnoses of ADHD and Oppositional Defiant Disorder (ODD). Jill was also his teacher. At the time of the interview with Jill, timeout had been in place for Ian for 2 weeks.

Charley was a 5-year-old boy who had a diagnosis of Sensory Processing Disorder. Charley's teachers were Paula and Dorothy. Paula had been teaching for 17 years, and Dorothy

had been teaching for 23 years. At the time of the interview with Paula and Dorothy, timeout had been in place for Charley for approximately 2 years.

Setting. Willis, Sonny, Keith, and Ian attended an alternative-education center that served elementary-aged students who engaged in severe challenging behavior. The center had three classrooms, access to a gymnasium, and access to a playground, and it was part of the county's public-school system. Charley attended a private preschool that served infants to kindergarten-aged children. The preschool had several classrooms, but observations of Charley only took place in Paula and Dorothy's classroom. The preschool was not part of the county's public-school system.

We observed the entire duration of regularly scheduled recesses for Willis, Sonny, Keith, and Ian. For Ian, we also collected data during the entire duration of trade-in times. Trade-in times were periods of the day during which Ian was allowed to exchange tokens for preferred activities. For Charley, we collected data during the entire duration of free-choice play or for 120 min, whichever came first. Free-choice play was a period of time during which Charley was allowed to engage in child-directed play with various activities.

Operational Definitions. The specific definitions of problem behavior for each student depended on the behavior reported by teachers during the interview and included in the BIP. Willis engaged in aggression defined as actual or attempted forceful contact with another person (including throwing objects at someone). Sonny engaged in aggression defined as actual or attempted hitting (with open hand or closed fist), kicking, pinching, biting, scratching, spitting on or grabbing another person, or throwing objects within 1 ft of another person. Keith engaged in peer aggression defined as actual or attempted forceful contact between some part of his body

and a peer (i.e., hitting, head butting, biting, or hair pulling). This also included spitting on a peer and throwing objects within 1 ft of a peer.

Ian engaged in aggression, property destruction, and negative peer interactions. Aggression was defined as actual or attempted forceful contact with another person including throwing objects at someone. Property destruction was defined as ripping, swiping, throwing, banging, or kicking materials, toppling furniture, or breaking materials. Negative peer interactions were defined as teasing peers (including name calling), telling peers what to do, or instructing peers to engage in problem behavior.

Charley engaged in aggression, property destruction, loud vocalizations, and language. Aggression was defined as actual or attempted forceful contact with another person (i.e., hitting, spitting, scratching, kicking, pushing, rough play), puffing chest out while grimacing and leaning into a peer's space, and throwing objects at another person. Property destruction was defined as throwing, kicking, tossing, ripping, sweeping, or stepping on materials or destroying a peer's creation. This definition excluded kneeling, sitting, or crawling on toys. Loud vocals were defined as vocalizations above conversation volume. Language was defined as swearing or pretending to shoot guns.

Timeout Procedures. Each participant had an individualized timeout procedure (see procedural component column in tables in Appendix B). Willis was required to sit for 1 min following instances of aggression (Table B1). Sonny was required to sit for 1 min without problem behavior following the first instance of aggression. Following the second instance of aggression, he was required to sit until the end of recess (Table B2). Keith was required to sit for 1 min without problem behavior following peer aggression (Table B3). Charley was required to sit for 10 s without problem behavior following aggression, property destruction, loud

vocalizations, and language (Table B4). Ian was required to sit for 1 min without problem behavior following aggression, property destruction, and negative peer interactions (Table B5).

Some components of each timeout procedure differed across students. Thus, Table B6 compares each student's procedure to best-practice recommendations about what components should be included in timeout. Each procedure included most of the components considered best practice (e.g., brief timeout duration, continuous timeout schedule, no attention during timeout, etc.). However, Sonny's procedure did not specify that attention should not be provided during timeout. Keith and Ian's procedure did not specify that access to items should not be provided during timeout. Charley's procedure did not specify that attention or access to items should not be provided during timeout.

Measurement and Data Collection. Trained observers collected continuous data using laptops with data-collection software that allowed the data collector to record responses and events in real time. The program produced time-stamped data records of response and event sequences. Each student was observed twice per week on average (range, 1-6 observations per week). Observations took place during recess or play and lasted for 28.72 min on average (range, 6.35-120 min). We collected data on the problem behavior reported by the teachers for a minimum of five school days and until behavior was either stable or increasing.

We also collected data on several environmental events, including warnings about timeout, instructions to go to timeout, physical guidance to timeout, the occurrence of timeout, adult and peer attention during timeout, and access to materials during timeout. See Table 1 for operational definitions. We collected data on these events to determine how often teachers implemented timeout following problem behavior. These data also provided information about how closely the timeout procedures aligned with what the teachers described in the interviews

and the procedure outlined in the BIPs, and whether the student received access to any potential reinforcers (e.g., toys, attention, etc.) during timeout. When more than one teacher implemented timeout (i.e., Charley), we collected data as if they were one person; we did not collect data on which teacher implemented each component.

Interobserver Agreement (IOA). A computer program was used to calculate the IOA scores using an interval-by-interval calculation. Each observer's data was divided into 10-s intervals. For each 10-s interval, the program calculated IOA for each response by dividing the smaller count by the larger count and converting to a percentage. For example, if Observer 1 scored teasing once and Observer 2 scored teasing twice during the first interval, the IOA for that interval would be (1/2)*100, yielding a score of 50%. If both observers did not score a response in the same interval, the IOA for that interval was 100%; both observers agreed that behavior did not occur.

After calculating IOA between the two observers for each interval, the program averaged the scores across intervals in the observation for each response to produce an average IOA score for that response during that observation. The program repeated this process for every behavior or event scored during the observation. Data collectors were previously trained by practicing computerized data collection on at least two different research projects until IOA was at least 80% across three consecutive sessions.

A trained, independent data collector collected data for 33% of observations for Willis, 86% of observations for Sonny, 60% of observations for Keith, 44% of observations for Ian, and 40% of observations for Charley. Average IOA across all behaviors and events was 98% for Willis, 99% for Sonny, 99% for Keith, 100% for Ian, and 99% for Charley. Tables 2 and 3

present average IOA percentages and ranges for all students across all target behaviors and events.

Data Analysis. We calculated omission and commission integrity percentages for each teacher. We also analyzed how closely the timeout procedures implemented aligned with what each teacher described during interviews and what was outlined in the BIPs. In addition, we analyzed whether attention or access to items was provided during timeout and the rates of problem behavior for each observation.

Omission-Integrity Calculation. Omission errors were said to occur when the teacher did not implement timeout (student within 1 ft of the timeout area) within 120 s of an instance of the targeted problem behavior. To calculate omission integrity, we first found each instance of targeted problem behavior in the data records. Then, we added 120 s to the timestamp of the instance of problem behavior. If the student entered the timeout area within the 120 s following the problem behavior, the timeout was counted as a correct implementation. For example, if peer aggression occurred at Second 10 in the observation, we looked to see if an instance of timeout was scored during Seconds 11 through 130. If timeout was scored within this window, we considered it a correct implementation. We repeated this process for every instance of problem behavior in the data records. If multiple instances of problem behavior occurred in a burst (instances within 2 s of each other), we still analyzed each individual instance of behavior. However, we also analyzed instances of behavior that occurred in a burst as if it were one instance of behavior, producing a second integrity percentage.

To determine the integrity percentage, we divided the instances (or bursts) of problem behavior followed by timeout within 120 s by total instances (or bursts) of problem behavior, and

multiplied by 100. If timeout did not occur during a 120-s window following problem behavior, we considered it an omission error and did not include it in the numerator of the calculation.

Commission-Integrity Calculation. Commission errors were said to occur when the teacher implemented timeout, but a targeted response had not occurred in the previous 120 s. To calculate commission integrity, we first found each instance of timeout in the data records. Then, we subtracted 120 s from the timestamp of the instance of timeout. If an instance of problem behavior occurred during the 120 s prior to timeout, we counted the timeout as a correct implementation. For example, if timeout occurred at Second 160 in the observation, we looked to see if an instance of problem behavior occurred during Seconds 40 through 159. If problem behavior was scored within this window, we considered it a correct implementation. We repeated this process for every instance of timeout in the data records. If problem behavior did not occur during a 120-s window, we considered it a commission error and did not include it in the numerator of the calculation. To determine the integrity percentage, we divided the total instances of timeout preceded by problem behavior within the previous 120 s by the total instances of timeout, and multiplied by 100.

Treatment Integrity. To compare how closely teachers were following the timeout procedures, the researcher used the output of computerized data collection to complete a treatment-integrity checklist based on the timeout procedure described during interviews and outlined in the BIPs. The checklists listed each component of the procedure. Tables B1-B5 in Appendix B show each student's checklist.

We used the data outputs to score the proportion of opportunities to implement each component correctly across observations. Zeros were recorded when the teacher never implemented the component correctly, and ones were recorded when the teacher implemented

the component correctly. For example, to score whether instructions to timeout were correct during an observation, we first found each instance of targeted behavior. Then, we evaluated if an instruction was scored within 10 s following each instance of targeted behavior. If problem behavior occurred in a burst (instances within 2 s of each other), we evaluated if an instruction was scored during the burst or following the last instance of the burst. If there was an instruction scored following each instance of behavior (or during or following a burst of behavior), we scored that as a correct implementation (recorded as a 1). If an instruction never followed an instance (or during or following a burst) of targeted behavior, we scored that as incorrect (recorded as a 0). If there were multiple opportunities to implement a component, integrity would be represented by a number between 0 and 1. For example, if three of four bursts of problem behavior were followed by an instruction, integrity for that component during that observation would be (3/4) = .75. We scored a component as "not applicable" if there were no opportunities to implement that component. For example, if no targeted behavior occurred during the observation, we scored the instruction component as not applicable (N/A). The integrity for each component was calculated by averaging across observations. We summed the number of times the teacher implemented a component correctly and divided by the total number of opportunities to implement that component across all observations. That number was then multiplied by 100 to yield a percentage.

The integrity for each observation was calculated by averaging across components. We summed of proportion of opportunities to implement each component correctly and divided by total number of components in which there was at least one opportunity to implement, and multiplied by 100. For example, if during an observation, one of seven components was implemented correctly, we divided one by the total number of components in which there was at

least one opportunity to implement the component (seven), and multiplied by 100. Thus, the integrity for that observation would be (1/7)*100 = 14%. The average integrity across all observations was calculated by summing the integrity per observation and dividing by the total number of observations.

Additional Data. We used the data outputs to analyze additional descriptive data (latency to instruction, latency to timeout, timeout duration, and duration of attention or access to items) and rates of problem behavior for each participant. The latency to a timeout instruction was calculated by subtracting the time targeted behavior occurred from the time an instruction was given. If an instruction was not scored, then that latency to instruction was scored as not applicable. The latencies to an instruction were summed across observations and divided by the total number of latencies that could be calculated to yield an average. The latency to the timeout area was calculated by subtracting the time the instruction was given from the start of timeout. Then, the latencies to timeout were averaged across observations. The timeout duration (time spent in the timeout area) was calculated by subtracting when timeout started from when timeout ended; then, the timeout durations were averaged across observations.

We also used the data outputs to analyze the provision of attention (from adults or peers) and access to items. To calculate the total duration of timeout during which attention or access to items occurred, we summed the number of seconds that attention or access to items occurred during timeout across observations. For example, if the total duration of attention during timeout was 120 seconds during Observation 1 and 20 seconds during Observation 2, the total duration of attention that occurred during timeout would be 140 s (or 2 min and 20 s).

To calculate the rate of problem behavior for each observation, we first subtracted the duration of time spent in timeout from the total observation duration. Then, we subtracted the

instances of problem behavior that occurred during timeout from the total instances of problem behavior that occurred during the observation. Finally, we divided the total instances of problem behavior (excluding those that occurred during timeout) by the duration of the observation (minus time spent in timeout). Instances of problem behavior during timeout and timeout duration were excluded from this calculation because if the transition to timeout evoked problem behavior, the overall rates of problem behavior would be inflated and influence the interpretation of the procedure's effectiveness.

Results

We observed Willis for 2.29 hr across six school days. Targeted problem behavior occurred 38 times; timeout occurred once following a burst of behavior. The upper left graph of Figure 1 shows omission and commission integrity results for Willis. The first bar represents the omission integrity. Timeout occurred after only 2 of 38 total instances of the target response. Those two instances of aggression occurred within 2 s of each other. Thus, the omission-integrity calculation based on individual responses yielded 5% integrity (2 individual instances / 38 individual instances*100). The omission-integrity calculation based on bursts yielded 3% integrity (1 burst of behavior / 37 bursts of behavior*100). We calculated the omission-integrity percentage based upon individual instances prior to Experiment 2 and the omission-integrity percentage based upon bursts after Experiment 2. Thus, the omission-integrity based upon individual instances is the percentage shown in Willis's graph. The second bar represents the commission integrity. Karl, Willis's teacher, never implemented timeout in the absence of targeted responding, resulting in 100% commission integrity.

Table B1 shows data collected on each component of Willis's timeout procedure across observations. Karl either always implemented a component or never implemented a component.

His average integrity was 3% (range, 0-17% integrity). Overall, Karl never read the rules script, stated the rule that Willis violated, instructed Willis to timeout (he just physically guided him), had Willis sit for the correct duration of time (he only had him sit for 12 s consecutively), or physically guided Willis back to timeout when he left the area. However, Karl did not provide any attention or access to materials during timeout.

We observed Sonny for 1.67 hr across seven school days. Targeted problem behavior occurred 15 times; timeout occurred four times. The upper right graph in Figure 1 shows omission and commission integrity results for Sonny. Timeout occurred after only 3 of 15 total instances of the target response. Thus, omission integrity was 20%. No bursts of behavior occurred. Jill implemented timeout once for disrobing (not a targeted response) resulting in 75% commission integrity.

Table B2 shows data collected on each component of Sonny's timeout procedure across observations. Jill's average integrity was 24% (range, 0-54% integrity). Overall, Jill never instructed Sonny to approach her and never implemented the first timeout for the correct duration (always longer than 1 min). However, she occasionally instructed Sonny to sit and always physically guided him to timeout when necessary, restricted access to items during timeout, and implemented the second timeout for the correct duration.

We observed Keith for 1.65 hr across five school days. Targeted problem behavior occurred 18 times; timeout occurred twice. The bottom left graph in Figure 1 shows results for Keith. Timeout occurred after 2 of 18 total instances of the target response. Thus, omission integrity was 11%. No bursts of behavior occurred. Kelly and Cathy, Keith's teachers, never implemented timeout in the absence of targeted responding, resulting in 100% commission integrity.

Table B3 shows data collected on each component of Keith's timeout procedure. Kelly and Cathy's average integrity was 50% (range, 40-60% integrity). Overall, they never instructed Keith to sit or had him sit for the correct duration. He always sat for longer than 1 min. However, they always remembered not to comment directly on problem behavior and instructed him to rejoin recess.

We observed Ian for 2.58 hr across nine school days. Ian never engaged in targeted problem behavior nor did Jill implement timeout. Because we could not calculate omission or commission integrity, Ian's data are not pictured in Figure 1.

We observed Charley for 9.53 hr across five school days. Targeted problem behavior occurred 185 times, and timeout occurred five times. The bottom right graph in Figure 1 shows results for Charley. Timeout occurred after 8 of 185 individual target responses (or after 5 of 181 bursts of behavior); thus, his omission-integrity percentage based upon individual responses was 4% integrity ([8 / 185]*100). His integrity based upon bursts was 3% integrity ([5 / 181]*100). We calculated the omission-integrity percentage based upon individual instances and the omission-integrity percentage based upon bursts prior to Experiment 2. However, because we had only calculated Willis's omission integrity based upon individual instances prior to Experiment 2, we presented Charley's omission integrity based upon individual instances in his graph. Dorothy and Paula implemented timeout in the absence of targeted responding once for using a rude tone of voice, resulting in 80% commission integrity.

Table B4 shows data collected on each component of Charley's timeout procedure. Paula and Dorothy's average integrity was 10% (range, 0-43% integrity). Overall, they never had Charley sit for the correct duration of time. He always sat for longer than 10 s when he was calm. However, they occasionally gave Charley warnings about timeout and instructions to sit.

In addition, they always guided him to timeout when necessary and started counting to 10 when he told them he was ready (by saying, “1”).

Table 4 shows additional descriptive data (latency to instruction, latency to timeout, timeout duration, and duration of attention or access to items) collected for each participant. Across students, the average latency to an instruction to timeout was 14 s (range, 6-28 s), the average latency to the timeout area was 54 s (range, 7-129 s), and the average timeout duration (time spent in the timeout area) was 70 s (range, 36-109 s). The latency to timeout and timeout duration varied depending on each student’s compliance with the procedure. More compliance resulted in quicker transitions to timeout and ending timeout sooner. Across students, the average duration of attention during timeout was 4 s (range, 0-10 s), and the average duration of access to items during timeout was 3 s (range, 0-10 s). These data indicate that the teachers were able to refrain from providing attention and restrict access to items during a majority of timeouts with their students.

Figure 2 shows the rate of targeted problem behavior for each participant (the y-axes are scaled to be the same as the y-axes in Experiment 2 graphs). These data show how frequently problem behavior occurred with each teacher’s naturalistic integrity. The upper left graph of Figure 2 shows the rate of aggression across six observations for Willis. Rate of aggression was 0.27 per min on average (range, 0.04-0.43 per min). The upper right graph shows the rate of aggression across seven observations for Sonny. Rate of aggression was 0.17 per minute on average (range, 0.00-0.43 per min). The bottom left graph shows the rate of peer aggression across five observations for Keith. Rate of peer aggression was 0.27 per minute on average (range, 0.03-0.61 per min). We observed no targeted problem behavior for Ian. Thus, his data are not pictured in Figure 2. The bottom right graph shows the rate of targeted problem behavior

across 10 observations for Charley. Rate of targeted problem behavior was 0.32 per minute on average (range, 0.03-0.75 per min). Rates of targeted problem behavior were highly variable for all participants.

Experiment 1 Discussion

Experiment 1 provided evidence of reduced integrity during naturalistic implementation of timeout. Overall, omission integrity was lower than commission integrity. In other words, teachers rarely implemented timeout following targeted problem behavior (low omission integrity) but also rarely implemented timeout following non-targeted behavior (high commission integrity). In addition, several of the timeouts implemented did not align with the components listed in the students' Behavior Intervention Plans (BIPs). Karl omitted all but one component. He provided no attention or items during timeout. Jill, Kelly and Cathy, and Dorothy and Paula implemented several components inconsistently (e.g., timeout instructions, timeout warnings, and timeout duration). Jill and Dorothy and Paula each also made one commission error.

Low omission integrity may have been positively correlated to the size of the class or the number of behaviors targeted for timeout. Charley's teachers, Paula and Dorothy, had the lowest omission integrity and the most students (15-20 children) to manage during play. In addition, Paula and Dorothy were targeting four types of problem behavior (aggression, property destruction, loud vocalizations, and language) for timeout. Karl, Jill, and Kelly and Cathy's classrooms ranged from approximately one to six children, and they were only targeting one type of problem behavior (aggression) for Willis, Sonny, and Keith. Thus, it may have been more challenging for Paula and Dorothy to catch every instance of four different problem behaviors while trying to manage more students. Although low, Karl, Jill, and Kelly and Cathy's integrity

was higher than Paula and Dorothy's integrity. When deciding which responses to target, it may be necessary to target the most important behaviors with the timeout procedure and use other reinforcement-based or antecedent-based procedures to address more minor forms of problem behavior.

Another explanation for the low omission integrity is teacher experience. At least one of Keith and Sonny's teachers were Board Certified Behavior Analysts (BCBAs) and Willis and Charley's teachers were not. Behavior-analytic training, in part, focuses on the development of strong data collection skills (and therefore, observation skills) and tolerance for the occurrence problem behavior while implementing behavior-reduction interventions. Being trained as a BCBA may have given Jill and Kelly the skills to observe the behavior of multiple students simultaneously and still catch several instances of problem behavior that should result in timeout. In addition, Jill and Kelly may have developed stronger skills to continue timeout implementation in the face of problem behavior evoked by timeout. Problem behavior evoked by timeout may have been more aversive to Karl and Paula and Dorothy, resulting in lower integrity. Jill and Kelly and Cathy's integrity was approximately three times higher than Karl and Paula and Dorothy's integrity, on average. In addition, Karl and Paula and Dorothy's integrity was very similar (only a 0.94 difference in percentage points). Thus, it might be important to measure and improve (if necessary) observation skills and increase tolerance for problem behavior before teaching teachers to implement timeout. However, had Jill and Kelly and Cathy had larger classrooms similar to that of Paula and Dorothy, their integrity might have been more similar to Paula and Dorothy's integrity. Thus, the relation between classroom size and teacher experience should be investigated further in future research.

Table B6 compares each student's timeout procedure to best-practice timeout recommendations. Most students' procedures included best-practice components. However, a few students' procedures included some variation. For example, Sonny's timeout schedule was an FR 1, but the duration of timeout varied depending upon the number of target responses that occurred. He was required to sit for 1 min following the first instance of problem behavior and for the remainder of recess following the second instance. As another example, Willis's procedure included a description of rules to follow during recess. Our experiment focused on the evaluation of one best-practice component, timeout schedule. However, additional research is needed on the necessity of other best-practice components and implementation of best-practice components in the real world.

The aim of this experiment was to conduct naturalistic observations of the frequency of timeout. It is possible that low omission integrity and inconsistent BIP implementation were observed because teachers were implementing the necessary components with the level of integrity needed to maintain manageable rates of targeted problem behavior. In other words, it may not be necessary to implement timeout for every instance of behavior. The existing timeout literature demonstrates support for the use of intermittent timeout (e.g., Clark et al., 1973, Experiment 2; Donaldson & Vollmer, 2012; Jackson & Calhoun 1977). Thus, in Experiment 2, we experimentally compared the omission integrity with which the teachers were implementing (intermittent timeout) to no-intervention baselines and 100% integrity phases.

Experiment 2

Method

Experiment 2 evaluated impacts of omission errors on the effectiveness of timeout. We expected that timeout would be most effective when implemented with 100% integrity. We also expected that reduced omission integrity would degrade the efficacy of timeout.

Participants, Setting, and Materials. Willis, Sonny, Keith, and Charley from Experiment 1 also participated in Experiment 2. Ian did not participate because we did not observe any problem behavior during Experiment 1. Sessions occurred during recess at an alternative-education center (Willis, Sonny, and Keith) or during free-choice play at a preschool (Charley) and lasted for the duration of the play situation. An average of 4 sessions (range, 1-5 sessions) were conducted with each student per week. Necessary materials included timers, data sheets, and pencils.

Measurement and Data Collection. We collected data on the frequency of targeted problem behavior across all phases of the experiment. Problem behavior for each student was the same as in Experiment 1, except for Keith. After Experiment 1, his teachers changed his BIP to include all negative peer interactions in addition to peer aggression as criteria for timeout. The definition for peer aggression was the same as in Experiment 1. Negative peer interactions were defined as threatening a peer with physical harm or instructing a peer to engage in problem behavior. Teasing was defined as any instance of calling peers names, labeling peers with negative adjectives or nouns (e.g., "You are… [dumb, a baby, a fatty rat, stupid, annoying]."), and any instances of pointing and sneering at a peer (including directly calling a peer a swear word [e.g., "You are a m*****f***er"]).

We collected data by tallying the instances of problem behavior across 1-min intervals. Problem behavior data were collected in 1-min intervals so that researchers could simultaneously implement timeout and collect IOA data. Researchers were individuals who were familiar with the students but did not interact with the students daily. We used paper data sheets (see Appendix C for representative example) with columns to collect data on target behavior. It is important to note that the data sheet included in Appendix C is only an example; the targeted

problem behavior written on the data sheet differed based on the targeted problem behavior for each student.

For example, using the data sheet in Appendix C, the data collector would record a tally mark following each instance of problem behavior. All topographies of problem behavior targeted for timeout (e.g., aggression, property destruction, negative peer interactions, etc.) were scored under the same column. Each instance that occurred before a timeout instruction would be marked in the column labeled, "Problem Behavior (TI)." Each instance that occurred after a timeout instruction would be marked in the column labeled, "Problem Behavior (Post-Instruction)." Instances that occurred during timeout would be marked in the column labeled, "Problem Behavior (TO)."

At the beginning of the session, each of the data collectors simultaneously started a re-looping, electronic timer. To ensure that they all started at the same time, one data collector counted down by saying, "3-2-1-press start," and each data collector pressed start on the timer when the data collector said, "start." The re-looping, electronic timer signaled the end of each interval by both vibrating and making a sound (e.g., a beep or a ding). When the timer started, time began at 0 s. Then, the timer counted up until the data collector pressed stop at the end of session (i.e., teacher instructed class to line up or clean up). To ensure that they all ended at the same time, one data collector counted down by saying, "3-2-1-press stop," and each data collector pressed stop on the timer when the data collector said, "stop."

During the session, the data collectors looked at the electronic timer to record the timestamps of timeout instructions in the column labeled, "Timeout Instruction," the start of timeout in the column labeled, "TO Start Time," and the end of timeout in the column labeled, "TO End Time." These data were collected as timestamps so that we could calculate latencies to

timeout and timeout durations. To collect the timestamp data, the data collector wrote the session time at which an instruction to timeout was given, when the student entered timeout (was within 1 ft of the timeout area), and when the student exited timeout (no longer within 1 ft of the timeout area) by looking at the session time displayed on the electronic timer. The data collector wrote the time (e.g., 5:26) in the appropriate column and row on the data sheet. At the end of the session (i.e., after pressing stop), the data collector also wrote the time (e.g., 20:18) displayed on the electronic timer in the session duration section at the bottom of the data sheet.

Interobserver Agreement (IOA). Secondary data collectors were trained using instructions, modeling, role play, and feedback during training observations. Data collectors were required to practice data collection during play situations until IOA scores for all responses were 90% or greater across two consecutive observations. Data collection practice first involved the researcher reviewing the operational definitions and data sheet of the student. Then, the researcher and the data collector collected data simultaneously and independently for the duration of the session. After the session, the researcher and data collector reviewed data collected together. During the review, they compared their data and the researcher provided positive feedback when data sheets matched and corrective feedback when the data were collected differently. Only observations in which at least one observer scored at least one instance of problem behavior during the session were used to determine when a data collector met the training criterion.

Agreement scores for problem behavior were calculated using interval-by-interval agreement. The smaller count of problem behavior was divided by the larger count of problem behavior for each 1-min interval. Then, those results for each interval were summed, divided by the total number of intervals, and multiplied by 100. Agreement scores for time-based measures

(timeout latency and timeout duration) were calculated using total agreement. First, we summed the latencies to timeout for the entire observation. Next, we summed the timeout durations for the entire observation. Then, we compared the duration each data collector recorded for the observation. The smaller duration was divided by the larger duration for each observation and multiplied by 100. After calculating IOA between the two observers for each session, we averaged the scores across sessions for each student. Interobserver agreement (see Table 5) was collected on 57% of sessions for Willis, 72% of sessions for Sonny, 62% of sessions for Keith, and 73% of sessions for Charley. On average, IOA was 94% (56%-100%) for Willis, 95% (11%-100%) for Sonny, 92% (38%-100%) for Keith, and 97% (90%-100%) for Charley.

Procedural Fidelity. A second observer collected procedural fidelity data on the researcher's behavior. Procedural fidelity was calculated as the number of procedural components implemented correctly divided by the total possible procedural components. Procedural components are listed in Tables B1-B7. Procedural fidelity was collected during 74% of sessions for Willis, 70% of sessions for Sonny, 68% of sessions for Keith, and 47% of sessions for Charley. On average, fidelity was 99% (range, 78%-100%) for Willis, 98% (range, 33%-100%) for Sonny, 97% (range, 33%-100%) for Keith, and 100% for Charley. Procedural fidelity was low for some sessions because the researcher and the fidelity data collector sometimes disagreed that a response occurred. If the fidelity data collector thought that an instance of aggression occurred and the researcher did not (and therefore did not implement timeout), only a few components would be applicable (1. does not comment directly on problem behavior, 2. describe the rule violated, and 3. gives an instruction to sit). If the researcher only implemented one component (did not comment), integrity was (1/3)*100 or 33%.

Obtained treatment integrity values were calculated for each phase. Table 6 shows the obtained integrity levels for each phase. Obtained treatment integrity is represented as the ratio of problem behavior followed by timeout to total instances of problem behavior. Neither problem behavior that occurred during timeout nor timeout duration were used during this analysis. In Table 6, the first column lists each phase. The second column lists the programmed ratio during each phase. The third, fourth, fifth, and sixth columns list the obtained ratio during each phase for each student. During baseline phases, the programmed ratio was 0:1 meaning no timeouts followed problem behavior. Across students, no timeouts occurred following problem behavior. During 100% phases, the programmed ratio was 1:1 meaning that every instance of problem behavior was followed by timeout. Across students, timeout followed nearly all instances of problem behavior. During the 5% phase for Willis, the programmed ratio was 1:20 meaning that one in 20 instances of problem behavior was followed by timeout. During the 11% phase for Keith, the programmed ratio was 1:9 meaning that one in nine instances of problem behavior was followed by timeout. For both Willis and Keith, timeout followed few instances of problem behavior.

For each session, we calculated the latency to the timeout area by subtracting the instruction timestamp from the timeout-start timestamp for each timeout. For example, if the instruction to timeout occurred at Second 55 and timeout began at Second 58, the latency to timeout was 3 s (58-55 = 3). We averaged the latencies across sessions within each phase. We also calculated the duration of timeout by subtracting the start of timeout from the end of timeout for each timeout in each session. For example, if the timeout ended at 1:58 and timeout began at Second 58, the timeout duration was 1 min (1:58-0:58= 1:00). We averaged the durations across sessions within each phase.

Table 7 shows the average latencies to timeout and the average durations of timeout for each student across phases. No data are included for Charley because we did not implement timeout phases. During 100% integrity phases, the latency to timeout was 35 s, on average, across students (range, 19 – 57 s). The duration of timeout was 3 min, on average, across students (range, 1 – 5 min). For Sonny, during the modified 100% integrity phase, the latency to timeout was 40 s, on average, and the duration of timeout was 2 min, on average. During reduced-integrity phases, the latency to timeout was approximately 21 s, on average, across students (range, 13 – 30 s). The duration of timeout was approximately 1 min, on average, across students (range, 1 – 2 min). The duration of timeout was shorter during the reduced-integrity phases because participants were meeting the criteria to end timeout sooner.

Procedure. We used a reversal design to demonstrate experimental control. At least five sessions, but no more than 15 sessions, were conducted per phase. The decision to change phases was based on visual inspection of graphed data or meeting the 15-session maximum. We did not control the interactions between students and their peers. We allowed students to play freely. If the student's behavior escalated to potentially dangerous behavior, we redirected peers to ignore or play in a different location (e.g., walk away). This redirection strategy was common in the classrooms.

We only instructed teachers and staff to not implement timeout (i.e., end the student's participation in play by having them sit out). Teachers talked to and played with the students as usual. The researchers and data collectors interacted with the students by talking to them or assisting them if the student approached and initiated the interaction. In addition, only the researchers implemented timeout when programmed.

Baseline. During this phase, problem behavior did not result in any programmed consequences. Teachers and staff interacted with the students by playing with them, talking to them, and assisting them when requested. However, they did not implement timeout.

100% Integrity. During each session, we instructed the student to timeout following each instance of targeted problem behavior by saying, "[Student Name], go sit." We followed each student's timeout procedure as described in his BIP. Access to all materials was restricted, and we did not attend to the student during timeout.

We modified timeout procedures for two participants. Willis often engaged in high rates of aggression when abrupt changes occurred in his environment. His psychiatrist recommended that we inform Willis of changes in his treatment to help him manage his own behavior. Thus, we described the timeout contingency to Willis prior to the first session of the 100% integrity phase. We also continued to read a script of the rules prior to each recess (across all phases) so that he was aware of the expectations we had set for him. These scripts are in Appendix D. The timeout procedure that Sonny's teacher implemented during Experiment 1 was ineffective. Thus, we modified the procedure. The steps of the modified procedure are in Table B7.

Reduced Integrity. During reduced-integrity phases, timeouts were programmed to occur intermittently. We used the level of omission integrity with which teachers implemented timeout from Experiment 1 to determine the frequency of timeout. For Willis, an average of one in 20 instances of targeted behavior was programmed to result in timeout (5% integrity). For Keith, an average of one in nine instances of target behavior was programmed to result in timeout (11% integrity).

The experimenter used a list to determine which instances of problem behavior would result in timeout. The list was created in Excel using the RANDBTWN function. For example,

Willis's 5% integrity list was created by using the RANDBTWN function to select a number between 1 and 20. When a "1" was next on the list, the next instance of problem behavior resulted in timeout. When "1" was not next on the list, the problem behavior did not produce a timeout.

All other procedures (e.g., timeout duration, restricted access to items during timeout, etc.) of this phase were the same as the 100% integrity phase. For Willis, we reminded him that timeout would occur following problem behavior and still read the rules prior to each recess (see Appendix D for reminder script). We were unable to implement a reduced-integrity phase with Sonny because we were initially unable to consistently reduce problem behavior with timeout implemented with 100% integrity, and the school year ended before we could finish the experiment. We were also unable to implement reduced-integrity phases with Charley because he was engaging in low rates of problem behavior during our initial baseline. Timeout was not warranted at that point in the experiment because problem behavior was occurring infrequently.

Data Analysis. We graphed rates of problem behavior during time-in for each session across phases. Instances of problem behavior during timeout were excluded from this calculation because if the transition to timeout evoked problem behavior, rates of problem behavior would be inflated and influence the interpretation of the procedure's effectiveness. Rates of problem behavior were calculated for each play situation by first subtracting instances of problem behavior that occurred during timeout (if applicable) from the total instances of problem behavior for that session. Then, we divided that number (total instances of problem behavior during time-in) by the time-in duration (session duration minus timeout duration).

Results

The upper left graph in Figure 3 shows data for Willis. Sessions are along the x-axis, and rates of aggression per min are along the y-axis. The solid vertical lines represent phase changes. During baseline phases, the average rate of aggression was 0.37 per min (range, 0.00-1.76 per min). Aggression varied from session to session and generally increased across sessions. When timeout was implemented with 100% integrity, aggression reduced to a mean of 0.10 times per min (range, 0.06-0.13 per min). During the 5% phase, rates of aggression remained low, occurring 0.04 times per min on average (range, 0.00-0.12 per min). Aggression was more stable across sessions during the 100% and 5% phases than during the baseline phases.

After session 32, Willis changed classrooms and attended recess with different peers. Willis engaged in few instances of aggression with his new peers, despite the absence of timeout. Given that aggression occurred infrequently during baseline after he changed classrooms, we decided that timeout was no longer warranted and discontinued data collection.

The bottom left graph in Figure 3 shows the instances of aggression followed by timeout (open circles) and the instances of aggression not followed by timeout (crosses) across phases. These data show that during baselines, no aggression resulted in timeout (crosses). During the 100% phase, all instances of aggression resulted in timeout (circles). During the 5% phase, four instances of aggression occurred, and one instance resulted in timeout. Three instances did not result in timeout.

The upper right graph in Figure 3 shows data for Sonny. During baseline phases, the average rate of aggression was 0.30 times per min (range, 0.00-1.04 per min). Rates of aggression were extremely variable, particularly in the second baseline phase. We first implemented the teacher's timeout procedure from Experiment 1 with 100% integrity. During

this phase (sessions 7-21) aggression occurred 0.19 times per min on average (range, 0.00-0.71 per min). Given that aggression was still variable and not consistently reduced after 15 sessions, we used the modified timeout procedure described in the method (see Table B7).

After modifying the procedure for Sonny, aggression occurred 0.11 times per min on average (range, 0.00-0.45 per min). Although the range of aggression had reduced, aggression was still variable session-to-session. At the end of the school year (sessions 79-83), aggression was occurring 0.05 times per minute on average (range, 0.00-0.10 per min). However, the school year ended before we were able finish this phase. Thus, we were unable to examine the potential effects of reduced integrity on rates of aggression.

The bottom right graph in Figure 3 shows the instances of aggression followed by timeout (open circles) and the instances of aggression not followed by timeout (crosses) across phases. These data show that during baselines, no aggression resulted in timeout (crosses). During the 100% phases, a majority of aggression resulted in timeout (circles).

The upper left graph in Figure 4 shows data for Keith. During baseline phases, the average rate of negative peer interactions was 0.44 times per min (range, 0.00-2.69 per min). Rates of negative peer interactions varied considerably. When timeout was implemented with 100% integrity, negative peer interactions reduced to a mean of 0.09 times per min (range, 0.00-0.26 per min). Negative peer interactions were more stable during the 100% phases than baseline phases. Sessions 48-53 represent a baseline phase that we ended early following his teacher's request. Rates of negative peer interactions during that phase were 0.88 per min on average (range, 0.00-2.69 per min). In addition, high rates of negative peer interactions often continued in the classroom after recess. Because his teacher was concerned about the safety of staff, the other students, and Keith, she requested that the phase be terminated. During the 11%

phases, the average rate of negative peer interactions was 0.14 times per min (range, 0.00-1.02 per min). During the first two 11% phases, negative peer interactions remained low occurring 0.18 times per min on average (range, 0.00-1.08 per min). The range of negative peer interactions during the first two 11% phases may have been so wide because baseline phases were implemented immediately before the reduced-integrity phases. When we implemented the 100% integrity phase prior to the last 11% phase, rates and variability in negative peer interactions were similar to those during the 100% integrity phase (average, 0.07 per min; range 0.00-0.31 per min).

The bottom left graph in Figure 4 shows the instances of problem behavior followed by timeout (open circles) and the instances of problem behavior not followed by timeout (crosses) across phases. These data show that during baselines and 11% phases, several instances of problem behavior did not result in timeout (crosses). During the 100% phases, a majority of problem behavior resulted in timeout (circles).

The upper right graph in Figure 4 shows data for Charley. Rates of problem behavior during the initial baseline phase were 0.09 per min on average (range, 0.00-0.62 per min). Because problem behavior was occurring infrequently in the absence of timeout contingencies, we decided timeout was no longer warranted and discontinued data collection.

The bottom right graph in Figure 4 shows the instances of problem behavior followed by timeout (open circles) and the instances of problem behavior not followed by timeout (crosses) across phases. Only initial baseline data were collected for Charley before participation discontinuation. These data show that no problem behavior resulted in timeout (crosses).

Experiment 2 Discussion

For two participants (Keith and Willis), timeout was programmed to be implemented with a very low level of integrity (11% and 5%, respectively) and was still effective at suppressing problem behavior. Despite intermittent timeout implementation, the total instances of problem behavior decreased during reduced-integrity phases compared to baseline phases (bottom graphs of Figures 3 and 4). These results are not surprising given that intermittent timeout can be effective (e.g., Clark et al., 1973, Experiment 2; Donaldson & Vollmer, 2012, Jackson & Calhoun, 1977), and the schedule of timeout can be successfully thinned from a continuous schedule to a more intermittent schedule (e.g., Donaldson & Vollmer, 2012). However, in Clark et al. (1973, Experiment 2), Donaldson and Vollmer (2012), and Jackson and Calhoun (1977), intermittent timeout was still occurring on a relatively rich schedule (e.g., VR 4, VR 5, VR 8). In our experiment, leaner schedules of timeout were effective (e.g., random-ratio [RR] 9, RR 20).

The effectiveness of lean schedules of timeout for Keith and Willis may be in part due to the most recent timeout history. Keith may have engaged in more negative peer interactions during some sessions in the first two 11% phases than the last 11% phase because the first two 11% phases were preceded by baselines (no timeout was being implemented). Recall that during 11% phases, several instances of negative peer interactions were not followed by timeout. Thus, from the Keith's perspective, the 11% phases may have seemed more similar to the baseline phases than the 100% phases. When we implemented a 100% integrity phase prior to an 11% phase, we did not observe sessions during which Keith engaged in increased rates of negative peer interactions. Willis only completed one 100% phase and one 5% phase, and both were preceded by a baseline phase. Aggression occurred at low rates and with similar levels of

variability during the 100% phase and the 5% phase. However, for Willis, it is difficult to assess whether recent timeout history impacted rates of aggression similarly to Keith because the 5% phase was never preceded by a 100% phase. In addition, we did not observe the wide range in problem behavior with Willis that we observed with Keith. One important difference between their procedures is that we did not read any rules to Keith prior to recess. Perhaps reminding Willis what behavior would result in timeout contributed to continued suppression of aggression. Similar rates and variability may have been observed regardless of a recent history of no-timeout implementation (baseline) because of the rules component to his timeout procedure.

Rules might mitigate detrimental effects of reduced integrity. For Willis, the reduction in aggression during the 5% phase may have been partly due to the rules script read prior to each recess. Timeout implemented with 5% integrity continued to suppress aggression despite a recent history of poor implementation (baseline). The rules may have developed stimulus control over behavior after being paired with the 100% phase. Willis was told what would occur if he engaged in problem behavior, and when he engaged in problem behavior during recess, the researchers implemented those previously verbalized contingencies. Because the rule was predictive of environmental contingencies during the 100% phase, the rules alone (without the contingencies) may have suppressed behavior during the reduced-integrity phase. In addition, the contingencies continued to occur intermittently during the reduced-integrity phase. Although less reliable, the rules still predicted environmental contingencies some of the time. Because Willis had had a previous reinforcement history with following the rules (if you follow the rules, you avoid timeout) during the 100% phase, these effects may have carried over into the reduced-integrity phase. Unfortunately, we were unable to replicate Willis's findings within or across participants. Thus, more research on the effects of rules on timeout implementation is warranted.

For Sonny, we were unable to develop an effective timeout procedure. Rates of aggression were variable across sessions, regardless of whether timeout was being implemented with 100% or 0% integrity (baseline). We may not have been able to develop a successful timeout procedure because time-in during recess may not have been reinforcing. Although Sonny may have liked recess and may have verbalized that he wanted to participate in recess, we did not evaluate whether recess was a reinforcer. In addition, preventing Sonny from participating in recess for the programmed duration of time may not have been aversive and therefore, may not have been an effective punisher. Unfortunately, we were unable to further modify Sonny's timeout procedure or experimentally identify why timeout was not effective before he transitioned to another school.

General Discussion

The evaluation of omission errors during timeout could be considered an evaluation of intermittent timeout. The effectiveness of intermittent timeout has been demonstrated in the literature (e.g., Clark et al., 1973, Experiment 2; Donaldson & Vollmer, 2012; Jackson & Calhoun 1977). However, less is known about the effects of timeout implemented with reduced integrity. Although some research has evaluated the effects of reduced treatment integrity on timeout procedures (e.g., Northup et al. 1997; Rhymer et al. 2002; Taylor and Miller 1997, Experiment 1), none conduct both descriptive and experimental studies of reduced timeout integrity.

Descriptive timeout studies are needed to determine if implementation recommendations are being followed in real-world applications of timeout and to guide experimental timeout research. Thus, the current research had two specific aims. We first evaluated how treatment integrity failures occurred naturally in schools (Experiment 1), and then, we investigated how

treatment integrity failures impacted the effectiveness of timeout (Experiment 2). In Experiment 1, we found that problem behavior was rarely followed by timeout (low omission integrity). We also found that a majority of components included in each student's timeout procedure did follow best-practice recommendations, but the implementation of those best-practice components was poor. Teachers also omitted or incorrectly implemented several steps (e.g., timeout instruction, duration of timeout, etc.) of the students' timeout procedures.

Our studies focused on effects of overall omission integrity (failure to implement timeout). However, omission and commission errors during specific components of a timeout procedure could also impact its effectiveness. For example, would timeout remain effective if the student was guided to timeout without an instruction to go sit? Would timeout remain effective if the student was allowed to access items? Future research should evaluate which specific components of a timeout procedure (e.g., no attention during timeout) are necessary to suppress problem behavior.

In Experiment 2, we found that timeout implemented with extremely reduced integrity suppressed problem behavior for two participants. This adds further support for the effectiveness of intermittent timeout (Clark et al.,1973, Experiment 2; Donaldson & Vollmer, 2012, Jackson & Calhoun, 1977). Our findings, in combination with the existing support for intermittent timeout, are in conflict the common recommendation that timeout should be implemented on a continuous schedule. Given that timeout can be implemented on various intermittent schedules and suppress problem behavior, future research should evaluate when timeout should be implemented on a continuous schedule and when timeout should be implemented on an intermittent schedule.

For Keith, timeout implemented at reduced integrity was more effective following timeout implemented with high integrity than following timeout implemented with extremely

poor integrity (baseline). Based upon these results, previous history with timeout may play a role in its effectiveness. The importance of a participant's history with a procedure has been demonstrated in the literature. For example, St. Peter Pipkin et al. (2010, Experiment 3) systematically evaluated effects of differential reinforcement of alternative behavior (DRA) implemented with 50% integrity when it followed baseline and when it followed DRA implemented with 100% integrity. The 50% integrity phase was more effective when it followed a DRA implemented with 100% integrity than when it followed a baseline phase. These results are similar to what we observed with Keith (11% integrity was more effective following the 100% phase than baseline phases). St. Peter Pipkin et al. (2010, Experiment 2) also found that ascending sequences of integrity phases (e.g., 20%, 60%, 80%) resulted in less suppression of problem behavior than did descending sequences (e.g., 80%, 60%, 20%). These findings are similar to those of Donaldson and Vollmer (2012), who thinned the timeout schedule in a descending sequence (e.g., FR 1 to VR 5) and maintained suppression of problem behavior.

Although implementation history may be an important factor in timeout effectiveness, few studies have systematically evaluated sequence effects using methodology similar to St. Peter Pipkin et al. (2010). Many timeout studies that use a reversal design alternate between baseline phases and timeout phases. The timeout phases often involve implementation of either continuous or intermittent timeout (e.g., Baseline-VR 8-Baseline-VR 4-Baseline-FR 1). Few studies include alternation between continuous timeout phases (i.e., FR 1 or 100% integrity) and intermittent timeout phases (e.g., Baseline-VR 8-Baseline-FR 1-VR 8). Comparisons between baseline-intermittent and 100%-intermittent phase sequences would evaluate the role of implementation history on timeout effectiveness, more clearly establishing if intermittent timeout is more effective following 100% integrity timeout than following baseline. If so, initial teacher

training may need to be mastery-based and time-intensive to ensure implementation of timeout with high integrity. Taking the time to well-train staff in the beginning may result in more robust treatment effects if integrity failures occur later. Our study only compared baseline-intermittent and 100%-intermittent sequences once with one participant, but data for this participant suggest that sequence may be important. Unfortunately, we were unable to replicate this sequence across participants due to low rates of problem behavior and participant attrition. Thus, future research should replicate those sequences and systematically compare other sequences (e.g., ascending versus descending integrity sequences).

The current research had several limitations. One major limitation was variability. To increase the likelihood that change in integrity level was the primary contributing factor to change in rates of problem behavior, we used a reversal design in Experiment 2 and often conducted phases for up to 15 days. However, only one recess occurred per day for students at the alternative-education center. Thus, we collected data during Experiment 2 for several months (five months per student, on average). In addition, these experiments were also conducted in schools. These two factors (data collection across several months and the dynamic school environment) restricted our control over several variables.

Uncontrolled variables that may have contributed to the variability we observed included play activity, staff present, and peers present. We allowed the activities and interactions with teachers, staff, and peers to vary. The games the students played often changed. In addition, the peers that were present varied. Peer absences could have impacted which games occurred during play. If the peer that always facilitated freeze tag, a game that can involve rough-and-tumble play, was absent, the students might have played independently (e.g., on the swings). For

participants who engaged in target behaviors towards others, peer absences may have reduced the likelihood that behavior targeted for timeout would occur.

We instructed teachers and staff to interact with peers as they normally would (except for implementing timeout in Experiment 2). However, teachers and staff sometimes needed to eat lunch or grade papers during play. These changes in teacher and staff activity across days may have impacted the amount of adult attention available, and the extent to which teachers responded to instances of target behavior during Experiment 1.

Another limitation was related to the definitions of problem behavior. The operational definitions for problem behavior were largely taken from the students' BIPs. The definitions may have made it difficult to catch every instance of behavior for some students. All participants' aggression definitions included "attempted" aggression. Sometimes it was difficult to determine when a student engaged in aggression on "purpose" or on "accident." The definitions of aggression captured several instances of pretend play (e.g., pretend fighting as Power Rangers). Although this type of play met the definition in the BIP, the teachers may not have implemented timeout following instances of pretend fighting that occurred within 1 ft of another person because they assumed that the students were not "trying" to engage in aggression. Teachers may have considered these instances of aggression as "accidents."

Difficulty objectively determining whether responses meet the operational definitions of problem behavior outlined in BIPs raises a few important questions. If teachers are likely to not record and subsequently fail to intervene upon problem behavior according to a student's BIP, it might be necessary to carefully consider social validity when creating BIPs. To remain as objective as possible, we scored instances of play fighting close to other people as aggression in both Experiments 1 and 2. However, implementing timeout for some of these may not be

socially valid to teachers even if they approve of an objective operational definition. Thus, when developing definitions for problem behavior that result in timeout, it may be important to include more specific operational definitions that clearly define which forms of play are and are not socially acceptable and roleplay use of that definition with the teachers. Ensuring that operational definitions are both objective and socially valid might increase teacher's omission integrity.

Although these experiments had several limitations, the results add to the existing timeout literature. Results from Experiment 1 provide a snapshot of naturalistic implementation of timeout in schools. The results suggest that when timeout is implemented, it is likely to be implemented intermittently. It is possible that teachers were implementing with the level of integrity necessary to reduce problem behavior to manageable rates in Experiment 1. Results from Experiment 2 add additional support for the effectiveness of intermittent timeout. For both Willis and Keith, timeout implemented with reduced integrity suppressed problem behavior to rates similar to timeout implemented with 100% integrity in Experiment 2. In addition, if there is a recent history of high-integrity timeout implementation, it may be possible to use intermittent timeout to maintain suppression of problem behavior.

Timeout is a behavior-reduction strategy that is effective in suppressing undesirable behavior across many populations and settings. In addition, it is a common strategy that caregivers, not just researchers, use. Additional research is needed on the variables (e.g., classroom size, number of target behaviors, specific components included in the procedure, teacher experience, history of implementation, detail of operational definitions, etc.) that may impact timeout's effectiveness. This additional research will add to the to the knowledge about how to best implement timeout so that its utility is maximized.

References

Alberto, P., Heflin, L. J., & Andrews, D. (2002). Use of the timeout ribbon procedure during community-based instruction. *Behavior Modification*, *26*(2), 297-311. doi:10.1177/0145445502026002008

Arkoosh, M. K., Derby, K. M., Wacker, D. P., Berg, W., McLaughlin, T. F., & Barretto, A. (2007). A descriptive evaluation of long-term treatment integrity. *Behavior Modification*, *31*(6), 880-895. doi:10.1177/0145445507302254

Barton, L. E., Brulle, A. R., & Repp, A. C. (1987). Effects of differential scheduling of timeout to reduce maladaptive responding. *Exceptional Children, 53*(4), 351-356. doi.org/10.1177/001440298705300410

Barton, E. S., Guess, D., Garcia, E., & Baer, D. M. (1970). Improvement of retardates' mealtime behaviors by timeout procedures using multiple baseline techniques. *Journal of Applied Behavior Analysis*, *3*(2), 77-84. doi:10.1901/jaba.1970.3-77

Bean, A. W., & Roberts, M. W. (1981). The effects of time-out release contingencies on changes in child noncompliance. *Journal of Abnormal Child Psychology, 9*(1), 95-105. doi:10.1007/BF00917860

Bostow, D. E., & Bailey, J. B. (1969). Modification of severe disruptive and aggressive behavior using brief timeout and reinforcement procedures. *Journal of Applied Behavior Analysis, 2*(1), 31-37. doi:10.1901/jaba.1969.2-31

Brantner, P., & Doherty, M. A. (1983). A review of timeout: A conceptual and methodological analysis. In S. Axelrod & J. Apsche (Eds.), The Effects of Punishment on Human Behavior (pp. 87-132). New York: Academic Press.

Burchard, J. D., & Barrera, F. (1972). An analysis of timeout and response cost in a programmed environment. *Journal of Applied Behavior Analysis, 5*(3), 271-282. doi:10.1901/jaba.1972.5-271

Calhoun, K. S., & Matherne, P. (1975). The effects of varying schedules of time-out on aggressive behavior of a retarded girl. *Journal of Behavior Therapy and Experimental Psychiatry*, *6*(2), 139-143. doi:10.1016/0005-7916(75)90039-7

Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis*, *46*(2), 379-394. doi:10.1002/jaba.49

Clark, H. B., Rowbury, T., Baer, A. M., & Baer, D. M. (1973). Timeout as a punishing stimulus in continuous and intermittent schedules. *Journal of Applied Behavior Analysis*, *6*(3), 443-455. doi:10.1901/jaba.1973.6-443

Crepsi, T. D. (1988). Effectiveness of time-out: A comparison of psychiatric, correctional and day-treatment programs. *Adolescence, 23*(92), 805-811. Retrieved from https://search-proquest-com.www.libproxy.wvu.edu/docview/195930468?accountid=2837

Cuenin, L. H., & Harris, K. R., (1986). Planning, implementing, and evaluating timeout interventions with exceptional students. *TEACHING Exceptional Children, 18*(4), 272-276. doi.org/10.1177/004005998601800408

DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis*, *44*(3), 611-615. doi: 10.1901/jaba.2011.44-611

Donaldson, J. M., & Vollmer, T. R. (2011). An evaluation and comparison of time-out procedures with and without release contingencies. *Journal of Applied Behavior Analysis*, *44*(4), 693-705. doi: 10.1901/jaba.2011.44-693

Donaldson, J. M., & Vollmer, T. R. (2012). A procedure for thinning the schedule of time-out. *Journal of Applied Behavior Analysis, 45*(3), 625-630. doi: 10.1901/jaba.2012.45-625

Donaldson, J. M., Vollmer, T. R., Yakich, T. M., & Van Camp, C. (2013). Effects of a reduced time-out interval on compliance with the time-out instruction. *Journal of Applied Behavior Analysis*, *46*(2), 369-378. doi:10.1002/jaba.40

Everette, G. E., Olmi, D. J., Edwards, R. P., Tingstrom, D. H., Sterling-Turner, H. E., & Christ, T. J. (2007). An empirical investigation of time-out with and without escape extinction to treat escape-maintained noncompliance. *Behavior Modification, 31*(4), 412-434. doi: 10.1177/0145445506297725

Fabiano, G. A., Pelham Jr., W. E., Manos, M. J., Gnagy, E. M., Chronis, A. M., Onyango, A. N., … Swain, S. (2004). An evaluation of three time-out procedures for children with attention deficit/hyperactivity disorder. *Behavior Therapy, 35*(3), 449-469. doi: 10.1016/S0005-7894(04)80027-3

Fee, V. E., Matson, J. L., & Manikam, R. (1990). A control group outcome study of a nonexclusionary time-out package to improve social skills with preschoolers. *Exceptionality*, *1*(2), 107-121. doi:10.1080/09362839009524747

Ferster, C. B., & Appel, J. B. (1961). Punishment of s^{Δ} responding in matching to sample by time out from positive reinforcement. *Journal of the Experimental Analysis of Behavior, 4*(1), 45-56. doi: 10.1901/jeab.1961.4-45

Foxx, R. M., & Shapiro, S. T. (1978). The timeout ribbon: A nonexclusionary timeout procedure. *Journal of Applied Behavior Analysis*, *11*(1), 125-136. doi:10.1901/jaba.1978.11-125

Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, *45*(2), 449-453. doi:10.1901/jaba.2012.45-449

Harris, K. R. (1985). Definitional, parametric, and procedural considerations in timeout interventions and research. *Exceptional Children*, *51*(4), 279-288.

Hobbs, S. A., & Forehand, R. (1977). Important parameters in the use of timeout with children: A re-examination. *Journal of Behavior Therapy and Experimental Psychiatry, 8*(4), 365-370. doi: 10.1016/0005-7916(77)90004-0

Hobbs, S. A., Forehand, R., & Murray, R. G. (1978). Effects of various durations of timeout on the noncompliant behavior of children. *Behavior Therapy*, *9*(4), 652-656. doi:10.1016/S0005-7894(78)80142-7

Huguenin, N. H. (1981). Nonexclusionary timeout: Maintenance of appropriate behavior across settings. *Applied Research in Mental Retardation, 2*(1), 55-67. doi: 10.1016/0270-3092(81)90006-0

Jackson, J. L., & Calhoun, K. S. (1977). Effects of two variable-ratio schedules of timeout: Changes in target and non-target behaviors. *Journal of Behavior Therapy and Experimental Psychiatry*, *8*(2), 195-199. doi:10.1016/0005-7916(77)90047-7

Kapalka, G. M., & Bryk, L. J. (2007). Two- to four- minute time-out is sufficient for young boys with ADHD. *Early Childhood Services, 1*(3), 181-188.

Kaufman, A. & Baron, A. (1968). Suppression of behavior by timeout punishment when suppression results in loss of positive reinforcement. *Journal of the Experimental Analysis of Behavior, 11*(5), 595-607. doi: 10.1901/jeab.1968.11-595

Mace, F. C., Page, T. J., Ivancic, M. T., & O'Brien, S. (1986). Effectiveness of brief time-out with and without contingent delay: A comparative analysis. *Journal of Applied Behavior Analysis*, *19*(1), 79-86. doi:10.1901/jaba.1986.19-79

Marlow, A. G., Tingstrom, D. H., Olmi, D. J., & Edwards, R. P. (1997). The effects of classroom-based time-in/time-out on compliance rates in children with speech/language disabilities. *Child & Family Behavior Therapy*, *19*(2), 1-15. doi:10.1300/J019v19n02_01

MacDonough, T. S., & Forehand, R. (1973). Response-contingent time out: Important parameters in behavior modification with children. *Journal of Behavior Therapy and Experimental Psychiatry*, *4*(3), 231-236. doi:10.1016/0005-7916(73)90079-7

McGuffin, P. W. (1991). The effect of timeout duration on frequency of aggression in hospitalized children with conduct disorders. *Behavioral Interventions, 6*(4), 279-288. doi: 10.1002/bin.2360060405

McKeegan, G. F., Estill, K., & Campbell, B. M. (1984). Brief report: Use of nonexclusionary timeout for the elimination of a stereotyped behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *15*(3), 261-264. doi:10.1016/0005-7916(84)90035-1

Northup, J., Fisher, W., Kahang, S. W., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities*, *9*(1), 1-16. doi:10.1023/A:1024984526008

Olmi, D. J., Sevier, R. C., & Nastasi, D. F. (1997). Time-in/time-out as a response to noncompliance and inappropriate behavior with children with developmental disabilities:

Two case studies. *Psychology in Schools, 34*(1), 31-39. doi: 10.1002/(SICI)1520-6807(199701)34:1<31::AID-PITS4>3.0.CO;2-Y

Pendergrass, V. E. (1971). Effects of length of time-out from positive reinforcement and schedule of application in suppression of aggressive behavior. *The Psychological Record*, *21*(1), 75-80. doi: 10.1007/BF03393992

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*(4), 477-492. doi:10.1901/jaba.1982.15-477

Pipkin, C. St. P., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis*, *43*(1), 47-70. doi:10.1901/jaba.2010.43-47

Porterfield, J. K., Herbert-Jackson, E., & Risley, T. R. (1976). Contingent observation: An effective and acceptable procedure for reducing disruptive behavior of young children in a group setting. *Journal of Applied Behavior Analysis, 9*(1), 55-64. doi:10.1901/jaba.1976.9-55

Rhymer, K. N., Evans-Hampton, T. N., McCurdy, M., & Watson, T. S. (2002). Effects of varying levels of treatment integrity on toddler aggressive behavior. *Special Services in the Schools, 18*(1/2), 75-82. doi: 10.1300/J008v18n01_05

Ritz, M., Noltemeyer, A., Davis, D., & Green, J. (2014). Behavior management in preschool classrooms: Insights revealed through systematic observation and interview. *Psychology in the Schools, 5*(2), 181-197. doi: 10.1002/pits.21744

Salend, S. J., & Gordon, B. D. (1987). A group-oriented timeout ribbon procedure. *Behavioral Disorders*, *12*(2), 131-137. doi: 10.1177/019874298701200205

Spitalnik, R., & Drabman, R. (1976). A classroom timeout procedure for retarded children. *Journal of Behavior Therapy and Experimental Psychiatry*, *7*(1), 17-21. doi:10.1016/0005-7916(76)90035-5

Taylor, J., & Miller, M. (1997). When timeout works some of the time: The importance of treatment integrity and functional assessment. *School Psychology Quarterly*, *12*(1), 4-22. doi:10.1037/h0088943

Warzak, W. J., & Floress, M. T. (2009). Time-out training without put-backs, spanks, or restraint: A brief report of deferred time-out. *Child & Family Behavior Therapy, 31*(2), 134-143. doi: 10.1080/07317100902910570

Warzak, W. J., Floress, M. T., Kellen, M., Kazmerski, J. S., & Chopko, S. (2012). Trends in time-out research: Are we focusing our efforts where our efforts are needed?. *The Behavior Therapist, 35*(2), 30-33.

White, A. G., & Bailey, J. S. (1990). Reducing disruptive behaviors of elementary physical education students with Sit and Watch. *Journal of Applied Behavior Analysis*, *23*(3), 353-359. doi:10.1901/jaba.1990.23-353

White, G. D., Nielsen, G., & Johnson, S. M. (1972). Timeout duration and the suppression of deviant behavior in children. *Journal of Applied Behavior Analysis*, *5*(2), 111-120. doi:10.1901/jaba.1972.5-111

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*(3), 369-373. doi:10.1901/jaba.2006.144-05

Wilson, C. C., Robertson, S. J., Herlong, L. H., & Haynes, S. N. (1979). Vicarious effects of time-out in the modification of aggression in the classroom. *Behavior Modification, 3*(1), 97-111. doi: 10.1177/014544557931006

Zimmerman, J., & Baydan, N. T. (1963). Punishment of s^{Δ} responding of humans in conditional matching to sample by time-out. *Journal of the Experimental Analysis of Behavior, 6*(4), 589–597. doi:10.1901/jeab.1963.6-589

Zimmerman, J., & Ferster, C. B. (1963). Intermittent punishment of s^{Δ} responding in matching to sample. *Journal of the Experimental Analysis of Behavior, 6*(3), 349-356. doi: 10.1901/jeab.1963.6-349

Table 1. *Operational definitions of environmental events that were the same across students for Experiment 1.*

Target Event	Operational Definition	Dimension
Warning about Timeout (Charley only)	Adult says, "If you [target behavior] again, you will have to go sit."	Frequency
Instruction to Timeout	Adult gives the student an instruction to go to timeout.	Frequency
Physical Guidance to Timeout	Adult physically guides student to the designated timeout area.	Duration
Timeout	Student is within 1 ft of the designated timeout area.	Duration
Adult Attention during Timeout	Adult talks to student or touches student.	Duration
Peer Attention during Timeout	Peer talks to student or touches student.	Duration
Access to Materials during Timeout	Student has access to an item (e.g., a toy) during timeout.	Duration

Table 2. Average *Interobserver Agreement (IOA) percentages and ranges for Willis, Sonny, and Keith for Experiment 1.*

Student	Peer Aggression	Aggression	Other Problem Behavior	Instruction to Timeout	Guide to Timeout
Willis	--	98 (95-100)	97 (93-100)	99 (99-100)	99 (97-100)
Sonny	--	99 (96-100)	99 (97-100)	99 (99-100)	99 (98-100)
Keith	99 (99-100)	99 (99-100)	100	99 (99-100)	99 (97-100)
	Timeout	Adult Attention	Peer Attention	Access	
Willis	93 (86-100)	100	99 (99-100)	100	
Sonny	99 (96-100)	99 (98-100)	100	100	
Keith	99 (98-100)	99 (98-100)	100	99 (98-100)	

Table 3. Average *Interobserver Agreement (IOA) percentages and ranges for Ian and Charley for Experiment 1.*

Student	Aggression	Property Destruction	Negative Peer Interactions	Loud Vocalizations	Language
Ian	100	100	100	--	--
Charley	99 (96-100)	98 (96-99)	--	98 (97-99)	100
	Other Problem Behavior	Warnings about Timeout	Instruction to Timeout	Guide to Timeout	Timeout
Ian	100	--	100	100	100
Charley	100	99 (99-100)	99 (99-100)	100	100
	Adult Attention	Peer Attention	Access		
Ian	100	100	100		
Charley	100	100	100		

Table 4. *Average (Avg.) latency to timeout instruction, latency to timeout area, timeout duration, total attention during timeout, and total access to items during timeout in seconds across students for Experiment 1.*

Student	Avg. Latency to Instruction	Avg. Latency to Timeout	Avg. Timeout Duration	Total Attention during Timeout	Total Access during Timeout
Willis	N/A	129	36	0	0
Sonny	6	7	109	2	0
Keith	28	28	82	2	10
Ian	N/A	N/A	N/A	N/A	N/A
Charley	7	53	54	10	0
Avg. across Students	14	54	70	4	3

Table 5. Average *Interobserver Agreement (IOA) percentages and ranges for Willis, Sonny, Keith, and Charley for Experiment 2.*

Student	Problem Behavior (Before Instruction)	Problem Behavior (After Instruction)	Problem Behavior (During Timeout)	Latency to Timeout	Timeout Duration
Willis	94 (80-100)	100	100	74 (56-88)	96 (95-98)
Sonny	96 (77-100)	99 (92-100)	99 (86-100)	80 (11-100)	97 (84-100)
Keith	92 (50-100)	99 (94-100)	97 (83-100)	72 (38-97)	97 (93-100)
Charley	97 (90-100)	--	--	--	--

Table 6. *Ratio of problem behavior followed by timeout (TO) to total instances of problem behavior per phase for Experiment 2.*

		Problem Behavior followed by TO : Total Problem Behavior			
Phase	Programmed Ratio	Willis	Keith	Sonny	Charley
Baseline	0:1	0:134	0:355	0:193	0:56
100%	1:1	5:5	27:29	20:20	--
Modified 100%	1:1	--	--	47:52	--
Reduced	1:20	1:4	--	--	--
	1:9	--	8:45	--	--

Table 7. *Average latency to timeout and duration of timeout in seconds per phase for Experiment 2.*

Phase	Willis		Sonny		Keith		Average across Students	
	Latency	Duration	Latency	Duration	Latency	Duration	Latency	Duration
100%	19	61	30	325	57	161	35	182
Modified 100%	--	--	40	125	--	--	--	--
Reduced	30	60	--	--	13	93	21	76

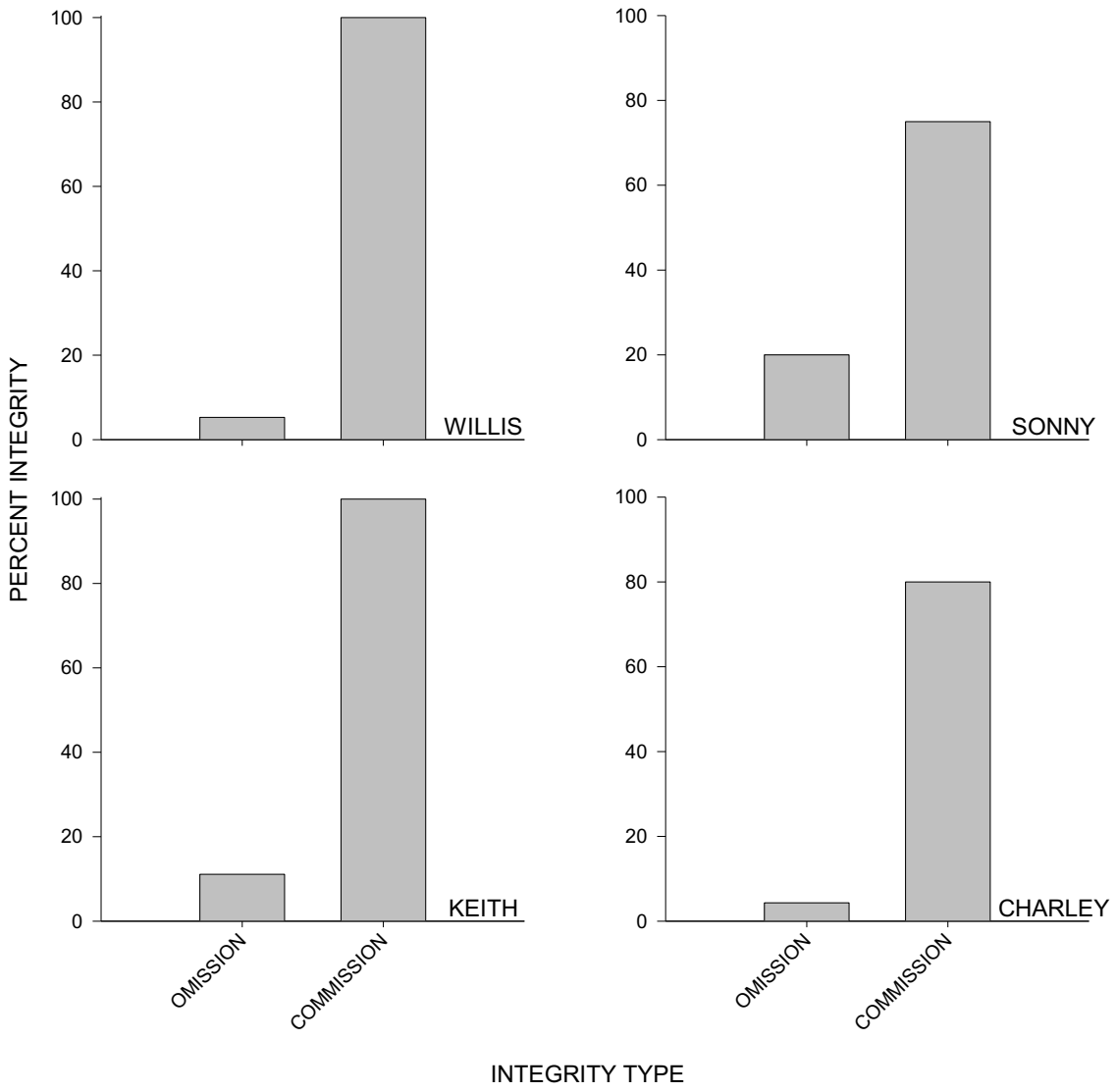

Figure 1. Percentage of omission and commission integrity across students in Experiment 1.

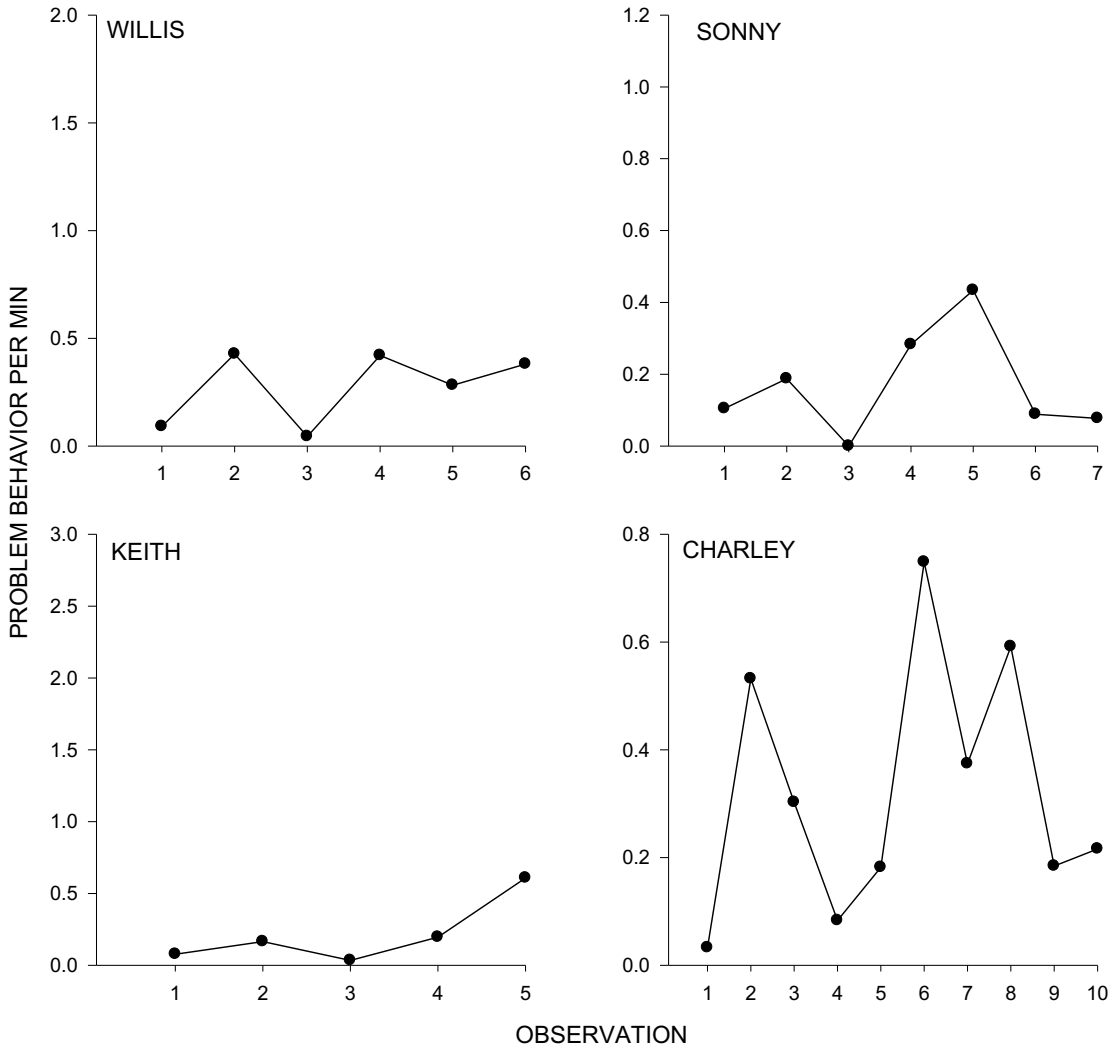

Figure 2. Targeted problem behavior per min across observations for each student in Experiment 1.

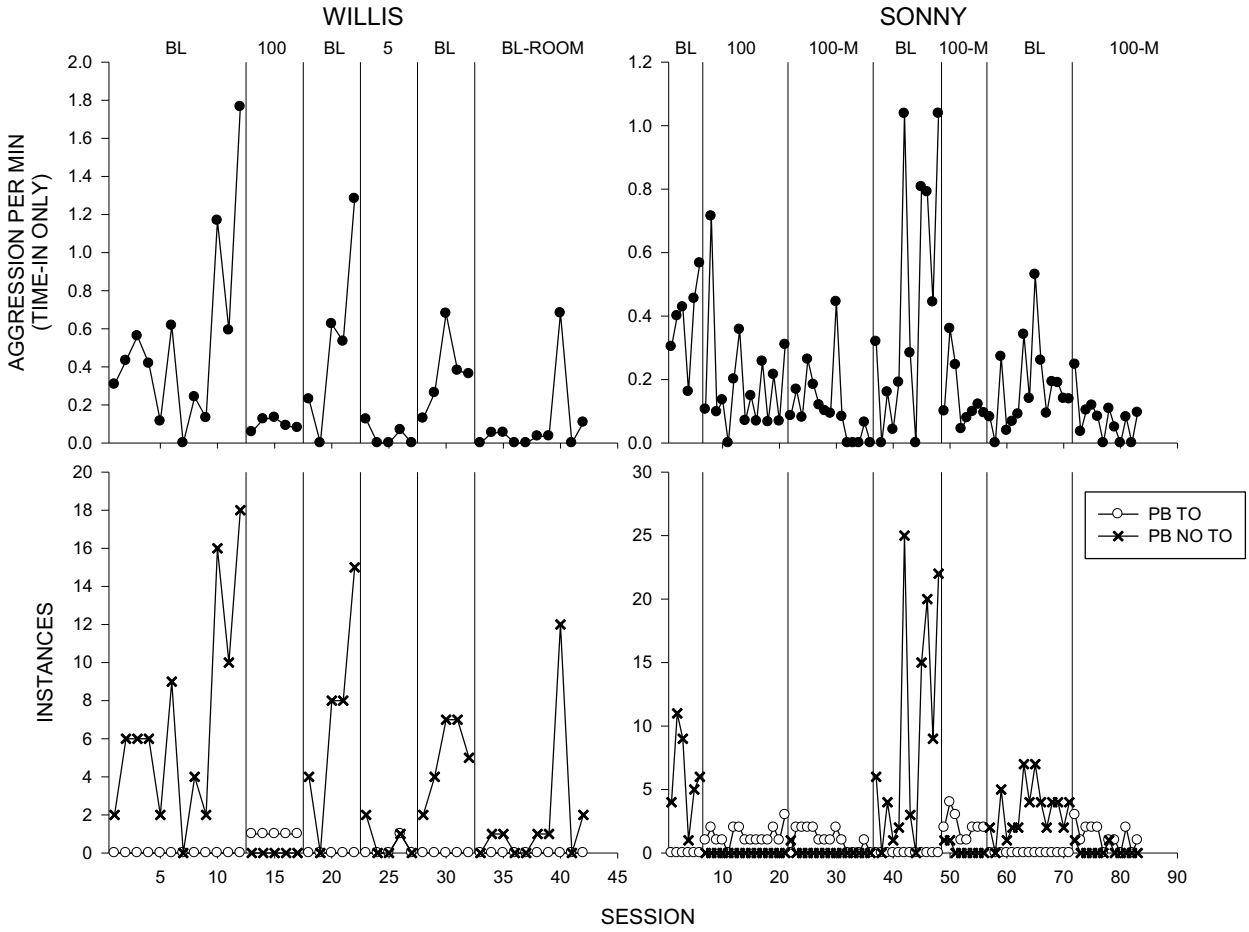

Figure 3. Top graphs are aggression per min and bottom graphs are instances of problem behavior followed by timeout (PB TO) and not followed by timeout (PB NO TO) for Willis and Sonny across sessions in Experiment 2.

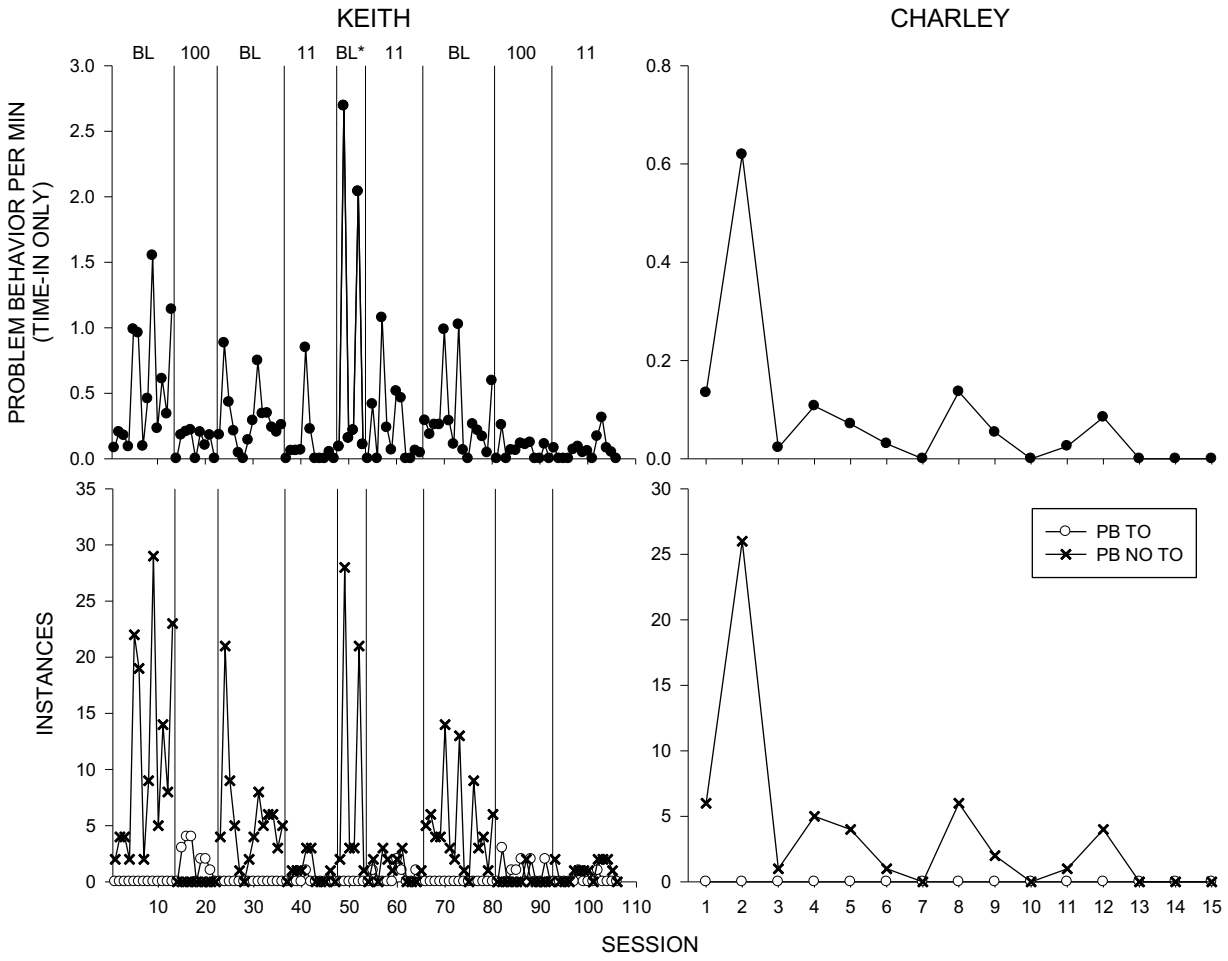

Figure 4. Top graphs are problem behavior per min and bottom graphs are instances of problem behavior followed by timeout (PB TO) and not followed by timeout (PB NO TO) for Keith and Charley across sessions in Experiment 2. For Keith, the baseline that was terminated early by his teacher's request is labeled with an asterisk (BL*). Only baseline data were collected for Charley.

Appendix A

Descriptive Evaluation of Behavior Management during Play
Teacher Interview

School:

Teacher Information:
Name: _____
Years teaching: _____

Student Information:
Name: _____
Date of Birth: _____
Diagnoses: _____
Existing Behavior Intervention Plan? YES NO
Is sitting out from play for misbehaving a part of that plan? YES NO

1. What behavior results in your student sitting out from play? What does it look like?

_____

_____

_____

_____

2. Does your student engage in any other problem behavior during play? If so, what does it look like?

_____

_____

_____

_____

3. How long have you been having your student sit out from play for misbehaving?

_____

_____

4. When your student misbehaves and needs to sit out, what do you do? Do you tell him/her to go sit out? If so, what's an example of what you would say?

_____

5. Do you provide any warnings before having your student sit out? If so, what's an example of what you would say?

6. What do you do if your student does not go sit on his/her own?

7. How long does your student have to sit?

8. What do you do if your student gets up before he/she is allowed to go play again?

9. Can your student go play immediately after time sitting is up? Do you give an explanation for why he/she had to sit?

10. What do you do, if anything, if your student is misbehaving when it is time for them to go play again?

Appendix B

Table B1. *Procedural integrity for Willis's timeout procedure for Experiment 1.*

Procedural Component	Observation 1	2	3	4	5	6	Component Integrity
1. Read the rules script prior to recess.	0	0	0	N/A	0	0	0%
2. Stated the rule Willis violated (e.g., "That's not keeping safe hands.").	0	0	0	0	0	0	0%
3. Instructed Willis to go sit out for 1 min (e.g., "Go sit for 1 minute.").	0	0	0	0	0	0	0%
4. Physically guided Willis to the timeout area if he did not comply with the instruction to sit.	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5. Had Willis sit for 1 min regardless of any problem behavior that occurred during the timeout.	N/A	N/A	N/A	N/A	0	N/A	0%
6. Physically guided Willis back to the area if he left the timeout area without permission.	N/A	N/A	N/A	N/A	0	N/A	0%
7. Did not provide any attention or access to materials during timeout.	N/A	N/A	N/A	N/A	1	N/A	100%
Integrity per Observation	0%	0%	0%	0%	17%	0%	
Average Integrity	3%						

Table B2. *Procedural integrity for Sonny's original timeout procedure for Experiment 1.*

Procedural Component	Observation 1	2	3	4	5	6	7	Component Integrity
1. Told Sonny to come to you.	0	0	N/A	0	0	0	0	0%
2. Told Sonny to go sit (e.g., "You weren't safe. You owe me 1 minute.").	1	0	N/A	.25	0	0	0	13%
3. Provided minimal physical guidance to the bench or timeout area if Sonny did not comply with the directive within 15 s.	N/A	N/A	N/A	1	N/A	N/A	N/A	100%
4. Restricted access to items (e.g., toys; excludes rocks, leaves, sticks, etc.).	1	N/A	N/A	1	N/A	1	N/A	100%
5. Had Sonny sit for 1 min for the first timeout.	0	N/A	N/A	0	N/A	0	N/A	0%
6. Had Sonny sit for the rest of recess for subsequent timeouts.	N/A	N/A	N/A	N/A	N/A	1	N/A	100%
7. Provided minimal physical guidance back to the bench or timeout area if Sonny left the area without permission.	N/A	N/A	N/A	1	N/A	N/A	N/A	100%
Integrity per Observation	50%	0%	N/A	54%	0%	40%	0%	
Average Integrity	24%							

Table B3. *Procedural integrity for Keith's timeout procedure for Experiment 1.*

Procedural Component	Observation 1	Observation 2	Observation 3	Component Integrity
1. Did not comment on the behavior.	1	1	1	100%
2. Provided one directive to sit out.	0	0	0	0%
3. Provided flat praise for compliance with directive.	N/A	N/A	N/A	N/A
4. Provided minimal physical guidance to the bench or timeout area if Keith does not comply with the directive within 15 s.	N/A	N/A	N/A	N/A
5. Had Keith sit for 1 min.	0	0	N/A	0%
6. Returned Keith to recess if he did not engage in problem behavior during the entire minute.	1	1	N/A	100%
7. Reset the 1-min timer if Keith engaged in problem behavior during the minute.	N/A	N/A	N/A	N/A
8. Did not provide high-quality attention during time out.	0	1	N/A	50%
9. Did not answer Keith if he appropriately requested an explanation of why he had to sit out.	N/A	N/A	N/A	N/A
Integrity per Observation	40%	60%	50%	
Average Integrity	50%			

Table B4. *Procedural Integrity for Charley's timeout procedure for Experiment 1.*

	Observation										
Procedural Component	1	2	3	4	5	6	7	8	9	10	Component Integrity
1. Provided a warning about timeout.	0	0	.07	0	.08	.06	.05	0	.09	.05	4%
2. Instructed Charley to go to sit.	0	.09	0	0	0	0	0	0	0	.10	2%
3. Physically guided Charley to timeout area.	N/A	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	100%
4. Started counting to 10 when Charley said, "1."	N/A	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	100%
5. Physically guided Charley back to timeout area if left.	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6. Had Charley sit for 10 s without problem behavior.	N/A	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0%
Integrity per Observation	0%	42%	3%	0%	4%	3%	2%	0%	5%	43%	
Average Integrity	10%										

Table B5. *Procedural components for Ian's timeout procedure for Experiment 1.*

Procedural Component
1. Prompted Ian to go to timeout area.
2. Physically guided Ian to timeout area if he did not comply with the directive within 15 s.
3. Told Ian that he had to remain in timeout for 1 min while using a quiet voice and calm body before he could return to the activity.
4. Provided praise for first appropriate request.
5. Ignored additional requests.
6. Provided no other attention during timeout.
7. Had Ian sit for 1 min without problem behavior.

Table B6. *Comparison of best-practice timeout components and procedural components across participants.*

Timeout Component	Best Practice	Willis	Sonny	Student Keith	Charley	Ian
Timeout Instruction	Instruction to sit[a]	Instruction plus description of rules prior to recess	Instruction to approach plus instruction to sit	Instruction to sit	Instruction to sit	Instruction to sit
Timeout Schedule	FR 1[b]	FR 1	FR 1	FR 1	FR 1	FR 1
Timeout Duration	5 min or less[b]	1 min	First violation – 1 min; Second violation – rest of recess	1 min	10 s	1 min
Attention during Timeout	No attention[c]	No attention	Did not specify	No attention	Did not specify	No attention
Access to Items during Timeout	No access[c]	No access	No access	Did not specify	Did not specify	Did not specify

[a]MacDonough and Forehand (1973)

[b]Brantner and Doherty (1983)

[c]Harris (1985)

Table B7. *Procedural components for the Sonny's modified timeout procedure in Experiment 2.*

Procedural Component
1. Did not comment on aggression.
2. Told Sonny to go sit (e.g., "You weren't safe. You owe me 1 minute.").
3. Provided minimal physical guidance to the bench or timeout area if Sonny did not comply with the directive within 15 s or moved in any direction away from the timeout area.
4. Restricted access to items (e.g., toys; excludes rocks, leaves, sticks, etc.) prior to Sonny sitting in timeout and within 10 s during timeout.
5. Had Sonny sit for 1 min without any aggression.
6. Provided minimal physical guidance back to the bench or timeout area if Sonny left the area without permission.
7. If Sonny did not sit within 15 s after physical guidance to timeout area, stated, "I can't start your time until you sit," once every 15 s.
8. Redirected peers that approached Sonny.
9. Did not provide attention during timeout.
10. Told Sonny he could go play again after the timeout.

Appendix C

Date:_*1/1/17*__ Student: _*WS*_ Researcher: _*AF*_ Session:_*35*_ Phase: _*100%*_ DC: _*GM*_ IOA: ***Y***/N

Interval	**Problem Behavior (TI)**	**Problem Behavior (Post-Instruction)**	**Problem Behavior (TO)**	**Instruction Time**	**TO Start Time**	**TO End Time**
Min 0-1						
Min 1-2	***I***			***1:05***	***1:30***	
Min 2-3						***2:30***
Min 3-4						
Min 4-5						
Min 5-6						
Min 6-7						
Min 7-8						
Min 8-9	***I***	***II***		***8:40***		
Min 9-10					***9:15***	
Min 10-11						***10:17***
Min 11-12						
Min 12-13						
Min 13-14						
Min 14-15						
Min 15-16						
Min 16-17	***I***			***16:37***		
Min 17-18					***17:01***	
Min 18-19						***18:02***
Min 19-20						
Min 20-21						
Min 21-22						
Min 22-23						
Min 23-24						
Min 24-25						
Min 25-26						
Min 26-27						
Min 27-28						
Min 28-29						
Min 29-30						
Min 30-31						
Min 31-32						
Min 32-33						
Min 33-34						
Min 34-35						

Time-in Rate	*3/15:33 = .19*	*2/1:24 = 1.43*	*0*			
Total Rate	*(3+2+0)/20 = .25*					

Session Duration: *20:00*
Total Timeout Duration: *3:03*
Total Post-Instruction Duration (Timeout Start Time minus Instruction Time): *0:25+0:35+0:24 = 1:24*
Total Time-in Duration (Session Duration minus Timeout Duration minus Post-Instruction Time): *15:33*

Appendix D

(For Willis only)

Recess Rules (Before each Recess)

"Before we go to recess, we need to review our recess rules. Remember to follow directions right away, keep safe hands and feet, use nice words, and play with materials appropriately. Also, remember to stay (in the gym/in the play area). You should not leave the (gym/play area) for any reason without permission from an adult."

Experiment 2 Timeout Modification: TO Script (Before First 100% Session)

"It seems like you are having trouble keeping safe hands and feet during recess again. So, we are going to start having you sit out from recess for 1 minute if you do not have safe hands and feet. That means that if you do not have safe hands and feet or it looks like your hands and feet are about to be unsafe, an adult [list possible adults] will tell you to go sit for 1 minute. After you sit for 1 minute, you will be able to go play again. Do you have any questions?"

Experiment 2 Timeout Modification: TO Script "Reminder" (Before First 5% Session)

"Sometimes you have trouble keeping safe hands and feet during recess. Remember, you will sit out from recess for 1 minute if you do not have safe hands and feet. That means that if you do not have safe hands and feet or it looks like your hands and feet are about to be unsafe, an adult [list possible adults] will tell you to go sit for 1 minute. After you sit for 1 minute, you will be able to go play again. Do you have any questions?"