Can a Brief Online Intervention Change Low-Income Caregivers’ Reported Use of Spanking? A Randomized Controlled Trial

Hilary L. Richardson
Western Michigan University, hilarylr2020@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Marriage and Family Therapy and Counseling Commons, and the Psychology Commons

Recommended Citation
Richardson, Hilary L., "Can a Brief Online Intervention Change Low-Income Caregivers’ Reported Use of Spanking? A Randomized Controlled Trial" (2020). Dissertations. 3636. https://scholarworks.wmich.edu/dissertations/3636

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
CAN A BRIEF ONLINE INTERVENTION CHANGE LOW-INCOME CAREGIVERS’ REPORTED USE OF SPANKING? A RANDOMIZED CONTROLLED TRIAL

by

Hilary L. Richardson

A dissertation submitted to the Graduate College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology
Western Michigan University
August 2020

Doctoral Committee:

Amy Damashek, Ph.D., Chair
Scott Gaynor, Ph.D.
Galen Alessi, Ph.D.
Dilip R. Patel, M.D.
Spanking is commonly used by parents (64-94%) in the United States as a strategy for managing undesirable child behaviors. Research has found that the use of spanking is particularly high among young mothers, low-income parents, and African American families. Decades of literature on the use of spanking has identified abundant detrimental outcomes for children such as increased externalizing behaviors, decreased long-term compliance, and less guilt following misbehavior, as well as serious outcomes in adulthood such as depressed mood and alcohol/drug use. There is also a risk for spanking to escalate to physical abuse. Thus, safer, more effective discipline strategies are recommended by research experts as well as the American Academy of Pediatrics (American Academy of Pediatrics, 1998).

There are currently no evidence-based interventions aimed at reducing parents’ use of spanking, and there is a particular lack of interventions that are broadly accessible to parents across a variety of life circumstances. Recent studies have evaluated a brief, online intervention called Play Nicely and found preliminary evidence that it decreases parents’ favorable attitudes toward spanking. These studies have some methodological limitations, though, and no research has attempted to measure Play Nicely’s impact on parents’ actual use of spanking as a discipline strategy. Additionally, little research has assessed whether the intervention is perceived as being culturally appropriate to users across racial/ethnic groups, and no study has examined whether
the intervention is equally effective in changing attitudes/behavior among White caregivers and caregivers of other racial/ethnic groups.

The present study used a randomized controlled trial design to: examine Play Nicely’s impact on attitudes toward spanking; evaluate Play Nicely’s impact on caregivers’ reported use of spanking; examine whether there were differences in treatment effects on attitudes toward spanking and reported use of spanking between White caregivers and caregivers of color in the treatment group; and examine whether there were differences in perceptions of the intervention’s cultural sensitivity between White caregivers and caregivers of color in the treatment group.

Participants in the treatment group demonstrated significantly greater changes in attitudes toward spanking from pretest to follow-up than did the control group. There was not, however, a significant difference between conditions in caregivers’ reported use of spanking at follow-up. Within the treatment group, there were no significant differences in outcomes between White participants and participants of color; specifically, there were no significant differences in changes in attitudes toward spanking or reported use of spanking between these subgroups. Finally, there was no difference in reported perceptions of the intervention’s cultural acceptability between White participants and participants of color in the treatment group.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude for the many people who have provided support and guidance throughout six years of planning, funding, conducting, and analyzing this dissertation. Thanks to the members of my committee, Dr. Amy Damashek, Dr. Scott Gaynor, Dr. Galen Alessi, and Dr. Dilip Patel, for their thoughtful feedback that greatly enhanced this project. Thanks also to Dr. Seth J. Scholer at Vanderbilt University for providing me permission to evaluate Play Nicely, and to the faculty and staff at the WMed pediatric residency clinic for facilitating such an advantageous setting for the study’s data collection. I owe a significant debt of gratitude to my labmates, cohort, classmates, and undergraduate research assistants, who were willing to allocate valuable spare time to sharing their perspectives on the topics herein at each step of this project. I am also eternally grateful to my loved ones who, throughout this process, showed me patience and understanding, gave me strength and hope when mine faltered, and kept me fed. Finally, thanks to my dog, Ango, for helping me get through many late nights of work by keeping my feet (and heart) warm.

Hilary L. Richardson
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................................................................................. ii

LIST OF TABLES ............................................................................................................................................. vi

LIST OF FIGURES ........................................................................................................................................... vii

INTRODUCTION .............................................................................................................................................. 1
  Summary and Study Goals ........................................................................................................................... 11

METHODS ...................................................................................................................................................... 13
  Participants .................................................................................................................................................. 13
  Recruitment ............................................................................................................................................... 14
  Procedures ................................................................................................................................................ 17
    Randomization ........................................................................................................................................ 18
  Intervention .............................................................................................................................................. 19
    Treatment group .................................................................................................................................. 19
    Control group ...................................................................................................................................... 22
  Measures .................................................................................................................................................. 23
    Demographics ...................................................................................................................................... 23
    Incredible Years Parent Practices Interview ..................................................................................... 23
    Child Behavior Checklist for Ages 1½-5 ............................................................................................. 24
    Attitude Toward Spanking survey ....................................................................................................... 25
    Parental Responses to Child Misbehavior ....................................................................................... 25
    Multicultural Therapy Competency Inventory .................................................................................. 26
Table of Contents—Continued

RESULTS......................................................................................................................................26

Descriptive Statistics ............................................................................................................26

Target child externalizing behaviors ...........................................................................26
Positive parenting practices .........................................................................................26
Attitudes toward spanking ...........................................................................................27
Perceptions of the intervention’s cultural sensitivity ..................................................27
Compliance with intervention engagement .................................................................28

Preliminary Analyses ............................................................................................................28

Strengthening constructs .............................................................................................28
Standardization ............................................................................................................29
Outlying and missing data ...........................................................................................30
Baseline comparisons ..................................................................................................31

Primary Outcomes ................................................................................................................31

Aim 1: Compare changes in attitudes toward spanking between conditions ..............31
Aim 2: Compare changes in reported use of spanking between conditions ...............33
Aim 3A: Compare changes in attitudes toward spanking between POC and White participants in the treatment group .................................................................33
Aim 3B: Compare changes in reported use of spanking between POC and White participants in the treatment group .................................................................34
Aim 4: Compare perceptions of the intervention’s cultural sensitivity between POC and White participants in the treatment group .................................................................34

DISCUSSION............................................................................................................................35

Limitations.............................................................................................................................38
Table of Contents—Continued

Implications ........................................................................................................................................39

REFERENCES ......................................................................................................................................40

APPENDIX

A. IRB Approval Letter ..................................................................................................................54
LIST OF TABLES

1. Demographic characteristics reported by participants at baseline .............................................46

2. Selection of final models’ fit statistics for constructs ATS, CBCL, and MTCI .........................50

3. ATS data by group and subgroup means (and SDs) across timepoints, in terms of raw scores, revised raw scores, and standardized scores .................................................................51

4. PRCM6 responses in frequency (and %) by condition at pretest and follow-up ......................52
LIST OF FIGURES

1. Participant flow ..............................................................................................................................................16

2. Menu screen for the interactive portion of the Play Nicely intervention .........................................................20

3. Visual information presented in Play Nicely when “Spank Your Child” option is selected .................................................................................................................21

4. Path diagram of the final factor analysis model for the Attitudes toward spanking construct (as measured by the ATS Survey) ........................................................................................................47

5. Path diagram of the final factor analysis model for the Target child externalizing behaviors construct (as measured by the CBCL Survey) ......................................................................................................48

6. Path diagram of the final factor analysis model for the Perception of intervention’s cultural sensitivity construct (as measured by the MTCI Survey) ..................................................................................49
Introduction

Spanking is one of many behaviors that falls under the umbrella of corporal punishment (CP). “Corporal punishment” may include any form of punishment that involves the use of physical force on a person’s body to cause pain or discomfort, but not injury, for the purpose of changing their behavior (e.g., pinching, striking with an object such as a belt, slapping face/head). The term “spanking” typically refers to hitting a child’s hands or buttocks with an open hand (Child Welfare Information Gateway, 2014). These terms are often used interchangeably, though, which may be one of the reasons for the broad variability among studies’ reports of the prevalence of spanking in the United States (i.e., use of all forms of CP reported as “spanking,” or the use of spanking reported generally as “CP”). Additionally, some studies report the proportion of parents who have used CP/spanking while others report the proportion of children who have experienced CP/spanking.

Parents’ use of CP and spanking in the United States is a common practice. A study from 1999 reported that over one-third of children under age 1 had already experienced some form of CP, and 94% of parents of 4- to 5-year-old children had used some form of CP within the previous 12 months (Straus & Stewart, 1999). Regarding the prevalence of spanking in particular, one study found that, among parents of children ages 1.5 to 3 years old, 64% reported that they had ever spanked their child and 26% reported that they frequently spank their child (Regalado, Harvinder, Inkelas, Wissow, & Halfon, 2004). In a more recent, longitudinal study that followed a large cohort of children over several years, data collected during the cohort’s kindergarten year (mean age of 6.17) revealed that 80% of the children’s mothers had spanked their child at least once (Gershoff, Lansford, Sexton, Davis-Kean, & Sameroff, 2012). Regarding the chronicity of spanking among parents who use it, one study extrapolated mothers’ reports of
the number of times they spanked in the past week to yield an annual estimate, which indicated that children between 3-5 years old are spanked an average of 150 times per year (Giles-Sims, Straus, & Sugarman, 1995). Despite the discrepancies between these reports of prevalence, they consistently support the conclusion that the use of spanking is common among families in the United States.

Research has identified characteristics of the parent, child, and the family environment that are associated with the likelihood of using CP. Mothers use spanking at a higher rate than do fathers (Wolfner & Gelles, 1993). This discrepancy may be partially explained by traditional parenting roles in which mothers have more contact with the child(ren) than fathers (Dietz, 2000; Straus & Donnelly, 1993); thus, it is possible that a more current survey would reflect recent shifts in family dynamics wherein it has become common for both parents to be involved with childcare. However, the large number of matriarchal single-parent households may also contribute to this parental gender difference in the use of spanking (Giles-Sims et al., 1995; McLoyd, 1990). Spanking also varies with the gender of the child, such that a greater proportion of boys are spanked than girls (Giles-Sims et al., 1995). The use of spanking is less common among younger mothers (ages 25-29) than it is among older mothers (ages 30-34; Giles-Sims et al., 1995). In terms of child age, the use of spanking peaks in chronicity among parents of 2- to 4-year-old children and then substantially declines with children’s age (Straus & Stewart, 1999).

Regarding cultural influences, a large national survey of women indicated that the use of spanking is more common among mothers in rural areas than those in urban areas, and that the prevalence of spanking is much higher in the South than in other regions of the United States (Giles-Sims et al., 1995). Additionally, racial/ethnic background is related to the likelihood of using CP. A study examining a nationally representative sample of approximately 2,000 parents
from the 2000 National Survey of Early Childhood Health found that the proportion of Black mothers who reported frequent use of spanking (31%) was twice as high as the proportion of White mothers (16%; Regalado et al., 2004). Another large study found that Black mothers reported the most frequent use of spanking, and Hispanic mothers reported greater use of spanking than White and Asian mothers (Gershoff et al., 2012). One study controlled for socioeconomic status and neighborhood quality and found that these variables did not reduce the statistically significant relationship between family race and the use of CP (Grogan-Kaylor & Otis, 2007). One explanation for this racial discrepancy is that cultural norms influence individuals’ internalized beliefs, and, because the use of CP is more normative among African American families, it is perceived as a culturally acceptable (and even useful) practice (Straus, 2010; Taylor, Hamvas, & Paris, 2011). Other caregiver factors associated with higher prevalence and/or chronicity of spanking include frustration and low emotional well-being (Regalado et al., 2004), as well as lower income, lower overall socioeconomic status, and being an unmarried mother (Giles-Sims et al., 1995). Given these factors, some researchers have suggested that the associations between spanking and being a single parent, African American, and/or impoverished may indicate an underlying “stress theory.” In other words, experiences of discrimination and/or restricted economic opportunity present added stress to the inherently demanding role of being a parent, so higher rates of CP among parents in these groups may be explained by the increased stress often associated with these experiences (Giles-Sims et al., 1995).

Despite its prevalence, many parenting experts agree that spanking is not a good discipline option for a number of reasons. One important cause for concern is that CP is a common antecedent to child maltreatment. Research has found that physical punishment often
escalates to physical abuse, especially among highly stressed parents (Gershoff, 2002; Graziano, 1989; Pinderhughes et al., 2000). One study found that 30% of parents reported using corporal punishment that was so harsh that it inflicted severe pain or caused welts and bruises. Moreover, 35% of children in this study reported being hit with objects such as sticks, whips, paddles, and cords. Perhaps even more alarming is that this study was conducted with a sample of low-risk families (i.e., middle-class, intact, suburban, well-educated). Given the nature of the sample in this study, rates of escalation to abusive behavior may be even higher among parents who are experiencing more stressful circumstances (Graziano, Hamblen, & Plante, 1996).

In addition to the risk of escalation to physical abuse, spanking is not recommended because it models aggressive behavior as a solution to interpersonal conflict, and it is not effective for improving child behavior. Numerous studies have concluded that spanking is associated with subsequent increases in externalizing behaviors for children of all races and ethnic groups (Gershoff et al., 2012; Lau, Litrownik, Newton, Black, & Everson, 2006; Pardini, Fite, & Burke, 2008). A review of existing meta-analyses indicated that, while spanking can produce immediate compliance, it is also associated with decreases in children’s long-term compliance, feelings of guilt following misbehavior, and attempts to repair the situation after harming others (Gershoff, 2002). Furthermore, researchers have found a positive correlation between how much a person is spanked and the likelihood of spanking their own children, approving of spousal violence, and experiencing more anger and marital conflict as an adult (Straus, 1996). This suggests that normalizing the idea of “legitimate violence” (i.e., growing up with the understanding that spanking and other types of CP are legal and acceptable forms of violence), in addition to perpetuating the use of CP, may result in a kind of spillover wherein a person can more easily justify the use of “criminal violence” (e.g., intimate partner violence).
Moreover, a recent study examined adult outcomes of spanking alongside those of Adverse Childhood Experiences (ACEs), and concluded that being spanked accounted for model variance on outcomes including suicide attempts, high alcohol use, and the use of street drugs, above and beyond variance explained by physical and emotional abuse (Afifi et al., 2017). Two recent reviews of the current literature on physical punishment bolster previous studies’ findings on the links between CP and detrimental outcomes for children and further support the conclusion that other strategies should be used in favor of spanking (Gershoff et al., 2018; Grogan-Kaylor, Ma, & Graham-Bermann, 2018). Fortunately, research has shown that discipline strategies such as time-out and removal of privileges produce positive long-term outcomes regarding the reduction of undesirable behaviors and are not accompanied by serious potential risks, and thus the American Academy of Pediatrics encourages caregivers to utilize these methods and not to use spanking (American Academy of Pediatrics, 1998).

Given the negative outcomes of CP and its widespread use, interventions are needed to educate caregivers on alternative discipline options. Research has found that some intensive interventions that provide education on effective discipline, such as The Incredible Years and SafeCare, effectively reduce future engagement in child maltreatment (Damashek & Chaffin, 2012). Additionally, a study on the Effective Black Parenting Program (EBPP), a 15-session intervention geared toward culturally competent parent training for Black families, found a specific outcome of reduced use of spanking for parents in one of the two treatment cohorts (Myers et al., 1992). Interventions such as these, however, are time-intensive and designed to address a variety of issues other than discipline. Moreover, with the exception of EBPP, they are typically only accessed by parents engaging in maltreatment or who have one or more children displaying significant behavior problems. CP is prevalent across family circumstances, so
briefer, more universally accessible interventions are needed to reach a broader range of the population.

In preliminary attempts to explore strategies that may reduce parents’ use of CP, existing studies have tested methods for changing participants’ beliefs about spanking, as measured by the Attitude Toward Spanking (ATS) survey. Specifically, several studies have tested the interactive component of a brief online program called Play Nicely in pediatric primary care clinics. Primary care settings have been selected as recruitment sites for these studies because they provide access to diverse families with varied exposure to risk factors and parenting expertise. Additionally, 48% of parents report that they are most likely to seek discipline advice from a pediatrician (Taylor, Moeller, Hamvas, & Rice, 2012), making pediatric clinics an ideal setting for delivering CP interventions to a high volume of families.

Play Nicely is a free online intervention aimed at educating parents/caregivers on strategies for managing aggression among children ages 1-7 (Scholer & Goad, 2003). The portion of Play Nicely that has been examined in studies aimed at changing beliefs about spanking presents users with a hypothetical scenario in which a young child hits another child. Then the program displays a menu of 20 options for responding to the situation. Users are asked to select and view a given number of response options, and are given immediate feedback about each response they choose and an explanation of why the option may or may not be effective (Scholer, 2008). Play Nicely has been identified as an attractive intervention for discipline recommendations because the program is easy to disseminate; it is brief, free, easily accessible, engaging, and no training is required to administer or use it. Thus, it offers a promising solution for reaching a large population of caregivers. From a public health perspective, even a small
effect size would be an indication that Play Nicely can produce desirable outcomes for a portion of the people who view it, thereby making the brief program’s dissemination a worthwhile effort.

One study that has evaluated Play Nicely’s influence on attitudes toward spanking, conducted by Chavis and colleagues (2013), recruited a sample of 258 parents of children ages 6-24 months from a pediatric primary care clinic. Parents were randomly assigned to treatment or control conditions, such that control participants received a routine primary care appointment and treatment group participants were exposed to Play Nicely as part of the child’s medical appointment (Chavis et al., 2013). Treatment group participants in this study were instructed to view at least 4 of the 20 interactive options (i.e., viewing duration of 5-10 minutes). All participants (n=258) were asked to complete a questionnaire (i.e., the ATS survey) immediately following their clinic visit. Chavis and colleagues (2013) found that, immediately following intervention, median ATS scores in the treatment group were significantly lower (i.e., less favorable attitudes toward spanking) than those in the control group. This finding is promising, but it is limited by the narrow and young child-age criterion, brevity of exposure to the intervention, and lack of follow-up data. Additionally, because pretest scores were not collected, it is unclear whether there was an existing difference between the treatment and control groups’ attitudes toward spanking at baseline.

In a study conducted by Scholer and colleagues (2010), a randomized controlled trial design was used to evaluate the effects of Play Nicely on a sample of 96 parents with a child younger than age 7, recruited from either a pediatric primary care clinic or a preschool. Parents in this study (n=96) were randomized to treatment and control conditions, asked to complete the ATS survey at baseline, and those in the treatment group were instructed to view at least 8 of the 16 interactive options presented in an earlier version of the Play Nicely program (i.e., viewing
times ranged from 8-20 minutes). Between 1-8 months after intervention, 64 of the 96 participants (66.6%) were successfully contacted to repeat the ATS survey over the telephone. Researchers did not find a significant difference in ATS scores between treatment and control groups at follow-up. They did, however, detect a significant change from pretest to follow-up ATS scores within the treatment group, with no significant change for the control group (Scholer et al., 2010). This study’s strengths are that it included baseline data and a more substantial exposure to the intervention, while its weaknesses are that immediate intervention effects were not measured, and the timeframe for posttest data collection varied broadly across the sample. Overall, the design of this study provides more conclusive results – but less robust support – for Play Nicely’s impact on parents’ attitudes toward spanking, given that a between-groups difference was not detected.

In addition to testing Play Nicely’s impact on parents’ attitudes toward spanking, several studies have assessed whether viewing Play Nicely provided parents with specific, realistic changes they could make in their approach to discipline, as well as the degree to which they intended to make such changes. Due to logistical constraints on researchers’ ability to obtain objective measures of parental discipline, these studies provide a useful preliminary indication of how parents reportedly received the intervention. In the above described study by Scholer and colleagues (2010), participants in the treatment group (n=45) were surveyed immediately after the intervention about whether they planned to change anything about the way they respond to their child(ren)’s misbehavior. Parents who endorsed plans to change how they discipline were then asked what they intended to do more or less. About 89% of participants who viewed Play Nicely indicated that they planned to change how they respond to their child’s behavior, and 22% specifically indicated intention to spank less (Scholer et al., 2010). Another study presented
Play Nicely to 197 parents during their child’s pediatric primary care visit, who were then asked the same post-intervention questions used in the previous study. Results indicated that 65% of parents in this sample planned to change how they discipline, and 9.6% reported plans to spank less. Parents who indicated that they planned to do less spanking were additionally asked one open- and one closed-ended question about how/why Play Nicely affected their intention to use spanking less. Responses yielded themes including, “I learned that spanking is not recommended,” and, “I learned alternative discipline strategies” (Hudnut-Beumler, Smith, & Scholer, 2018). These findings provide preliminary support for Play Nicely’s application as a parenting intervention that promotes effective, nonviolent discipline strategies, but there are important limitations to the existing data.

Previous studies examining the program’s impact on attitudes toward spanking instructed participants to view 4 or 8 options from the Play Nicely menu (i.e., 5-15 minutes of treatment), resulting in low and inconsistent treatment dosing. This very brief exposure to the intervention, in combination with methodological limitations (i.e., lack of pretest data, immediate posttest and/or more defined follow-up data collection), leaves ambiguity about the implications of the treatment effects that were detected. Preliminary data collected for this author’s thesis project, using a randomized controlled trial design in a pediatric primary care clinic, were analyzed for a sample of 52 participants. Participants completed a pretest and posttest measure on attitudes toward spanking and a posttest measure on perception of the program’s cultural sensitivity immediately following the intervention. A between-group effect size indicated a small difference in attitudes toward spanking at posttest ($d = 0.20$); however, the sample size of 52 participants did not provide sufficient power to detect a significant difference between groups on either
measure (Richardson & Damashek, 2019). Conducting the same procedures with a larger sample may reveal a statistically significant difference between groups.

Additional research is needed to address previous studies’ shortcomings, including the degree of exposure to the intervention, rigor of methods employed, and examination of possible contributing and/or confounding factors (i.e., sample income or cultural identities), in order to clarify Play Nicely’s impact on parents’ attitudes toward spanking. Furthermore, beyond the examination of parent’s beliefs about spanking, there are no known studies that have evaluated whether Play Nicely (or any other discipline-specific intervention) can effectively decrease parents’ use of spanking.

Another gap in the literature on CP interventions and research on Play Nicely is that research has not adequately attended to cultural differences in attitudes toward spanking. As mentioned above, this author’s thesis project assessed participants’ perceptions of Play Nicely’s cultural sensitivity and found that the treatment group rated it highly, but the sample was too small for any difference in ratings between White participants and participants of color to be detected (Richardson & Damashek, 2019). To this author’s knowledge, only one other study has focused on this issue with regard to Play Nicely. Play Nicely was presented to a racially/ethnically diverse sample of parents (n=197), and their perceptions of the program’s cultural sensitivity was compared (Smith, Hudnut-Beumler, & Scholer, 2017). Parents across racial groups reported that Play Nicely “addressed their family needs,” and “respected their family values and personal beliefs,” but the study did not evaluate the sample’s attitudes toward spanking. Another study on Play Nicely indicated that Black parents had a higher posttest median score on favorable attitudes toward spanking than did White parents (Chavis et al., 2013), but researchers have not evaluated whether the intervention is equally effective, in terms
of the degree of change in attitudes toward spanking, for parents of color (POC) and White parents. As indicated by Myers and colleagues in their article on the Effective Black Parenting Program, most parenting programs are designed for (and evaluated with) White, middle-class families, “and may be of questionable utility for many ethnic minority and low-income parents” (1992, p. 133). In recent years researchers have begun attending to the need for evaluating the cultural relevancy of existing evidence-based interventions and, when warranted, creating cultural adaptations thereof (Marsiglia & Booth, 2015). In the case of Play Nicely, though, given that its evidence base is still developing, an opportunity exists to give substantial consideration to important cultural factors as the literature evolves rather than circling back to them as an afterthought. Thus, it is important to examine whether Play Nicely is equally effective among parents of color and White parents, and to further examine perceptions of the intervention’s cultural sensitivity across families from diverse backgrounds.

**Summary and Study Goals**

The high prevalence of spanking and its associated negative outcomes call for an accessible, culturally competent intervention that specifically targets this problem. Play Nicely shows promise for fulfilling this need due to its ease of dissemination as well as preliminary research findings that support its ability to decrease parents’ favorable attitudes toward spanking. These findings are promising, given that research has indicated that attitudes toward spanking are significantly correlated with subsequent use of CP (Vittrup, Holden, & Buck, 2006). However, due to the more favorable attitudes toward and greater reported use of CP among parents of color (POC), it is important to examine whether Play Nicely is equally effective for caregivers of different racial and ethnic backgrounds. The goals of the current study were to use a randomized controlled trial design to: 1) Replicate previous studies’ examination of the effects of Play Nicely
on attitudes toward spanking, specifically among low-income caregivers of young children (ages 1-5); 2) evaluate Play Nicely’s impact on caregivers’ reported use of spanking; 3) examine whether there were differences in [A] the changes in attitudes toward spanking and [B] reported use of spanking between White caregivers and caregivers of color in the treatment group; and 4) examine whether there were differences in caregivers’ perceptions of the intervention’s cultural sensitivity between White caregivers and caregivers of color in the treatment group.

In an effort to clarify Play Nicely’s impact on attitudes toward spanking, this study was designed with a goal of addressing some of the methodological limitations of previous studies, including the collection of pretest, immediate posttest, and one-month follow-up data, a more significant exposure to the intervention, and the examination of variables that may impact participants’ treatment response (i.e., baseline use of positive parenting practices, perception of the program’s cultural sensitivity). The study built upon the student author’s thesis project in that it includes the analysis of one-month follow-up data on participants’ attitudes toward spanking, presents novel findings regarding Play Nicely’s impact on caregivers’ use of spanking, and achieved double the sample size to allow for more robust conclusions regarding the effects of the intervention. We hypothesized that: 1) Posttest and follow-up attitudes toward spanking would be lower in the treatment group than in the control group; 2) reported use of spanking at follow-up would be lower for the treatment group than the control group, and the effect size would demonstrate a small positive change for the treatment group; 3) attitudes toward spanking would be more favorable among participants of color compared to White participants at posttest and follow-up. Analyses regarding the differences between POC and White caregivers’ changes in attitudes toward spanking and their perceptions of the intervention’s cultural sensitivity were exploratory.
Methods

Participants

The sample consisted of 107 English-speaking, literate caregivers of at least one child aged 1-5 years old. The average age for the total sample (N=107) was 31.03 years, and the majority (92.5%) of the participants were female. Most participants were either married (43.0%) or never married (39.3%). Approximately half of the sample was White (50.5%) and 29.9% were African American. The rest were biracial (9.3%), American Indian or Alaskan Native (1.9%), Hispanic or Latino/a (2.8%), or Asian American (0.9%); five participants selected “Other” (4.7%). Approximately 66% of the sample reported education levels of “high school graduate” (25.2%) or “some college” (41.1%). Of the 100 participants who reported gross annual household income, 52.4% endorsed an annual household income of $24,999 or lower.

When compared with county-level (81.2% White) and national (76.3% White) population characteristics reported by the U.S. Census Bureau (2019), participants of color were overrepresented in this study’s sample. The education level reported by the sample is lower than average; an education level of “Bachelor’s degree or higher” is reported by 38.1% of individuals 25 and older in the county and 31.5% nationally. In addition, the U.S. Census Bureau indicates a median household income, in 2018 dollars, of $54,431 in this county (with 14.4% of persons in poverty) and $60,293 nationally (with 11.8% of persons in poverty)—thus indicating that the sample’s income is relatively quite low.

The majority of participants (85.0%) identified themselves as the mother of the target child; the rest were fathers (7.5%), grandparents (3.7%), or other (3.8%). Participants reported having an average of 2.56 children living in their home. Approximately half of the target children (53.3%) were male; two participants did not indicate their target child’s gender. The
mean age of participants’ target children was 37.42 months old. See Table 1 for summaries of demographics of the treatment and control groups and overall sample.

Recruitment

Subject recruitment took place at a pediatric primary care clinic. This site was chosen because parents frequently elicit recommendations on child discipline from pediatricians (Taylor et al., 2012). Moreover, previous studies on Play Nicely also occurred in pediatric clinics (Chavis et al., 2013; Scholer et al., 2010). The sample was recruited from the Western Michigan University School of Medicine’s pediatric residency training clinic. This particular clinic was chosen because it provides services to a low-income and a racially/ethnically diverse population. Specifically, a large majority of this clinic’s patients are insured through Medicaid, and the present study aimed to focus on a low-income population.

Individuals were ineligible for the study if they were not the parent or legal guardian of a child between ages 1-5, were not fluent in English, or were not literate. The language requirement was in place because the research team was unable to readily administer or answer questions about the consent form or questionnaires in other languages, and literacy was necessitated by the format of the intervention (i.e., reading about discipline options). The age criterion (parents/caregivers of children ages 1-5) was established because it is consistent with the population for whom Play Nicely was developed, and the instruments used in this study measure behaviors that may be displayed by children ages 1 to 5 as well as discipline strategies typically used for children in the same age range. Legal guardianship was included as an inclusionary criterion due to the knowledge that many children are brought to appointments at this clinic by foster parents, grandparents, or other relatives with either temporary or permanent child custody, and it was agreed that these individuals, in addition to biological parents, could
benefit from participation. Thus, the term “caregiver(s)” is used to describe the sample throughout this document because the sample included some individuals who were a primary guardian/caregiver, but not the biological parent, of a child between ages 1-5.

Upon checking in for their child’s appointment at the clinic, caregivers were given a flyer explaining the study. If interested, they wrote their name and phone number on the flyer, and checked boxes indicating they had a child between ages 1-5 and wanted to learn more about participating in the study. The flyer instructed interested caregivers to complete and return it to a research assistant (RA) in the waiting room or a nurse (who will then pass it on to the RAs). At the end of the child’s appointment, an RA greeted caregivers who submitted a flyer and led them to an examination room in the pediatric clinic that is reserved for research activities to briefly describe the study, answer any questions, and verbally confirm caregivers’ eligibility.

A total of 228 individuals completed a recruitment flyer indicating interest in study participation. In the event that a caregiver was interested in participating but unable to do so on the same day as their child’s appointment, RAs offered to schedule their participation for a subsequent day. Among potential participants who were contacted to schedule a day/time to return to the clinic to participate, 40 individuals withdrew interest (stating reasons such as limited availability and/or transportation, or no longer wanting to participate). An additional 72 individuals either could not be reached (i.e., did not answer/return phone calls) or did not attend their scheduled participation appointments (i.e., were removed as potential participants after two “no show” appointments). A total of 9 individuals were determined to be ineligible for the study because their child was younger than 1 year old, they did not currently have legal custody of their child, or they were a babysitter but not the parent or legal guardian of a child in the required age range. Thus, of the 228 individuals who indicated interest in participating, 121 people were
not enrolled. A total of 107 individuals were enrolled in the study, randomly assigned, and engaged in data collection procedures. There were 49 participants assigned to the treatment condition and 58 participants assigned to the control condition. Of the 107 participants, 7 individuals could not be reached for follow-up data collection (after 4+ attempts over the course of 2 weeks), resulting in a 6.54% overall dropout rate.

Figure 1. Participant flow
Procedures

On data collection days, there were at least two RAs available at the recruitment site, so all participants were given the option of having their child(ren) supervised by an RA in a separate room while they completed the research procedures. Once consent was obtained, participants completed a demographic measure. In the event that a participant had more than one child between ages 1-5, the RA used a procedure to randomly select the target child (i.e., the child they would answer questions about on study measures). Using standard playing cards, the RA shuffled a small deck that contained 2 cards representing each age (1 [Ace], 2, 3, 4, 5; hearts and spades). The RA asked the participant the ages of their children who are between 1-5 (e.g., 2-year-old and 4-year-old) then chose a card from the deck, repeating this until they pulled a card whose number matched the age of one of the potential target children (e.g., 2 of spades); that child was then identified as the target child. If a participant had twins in the target age range, we designated the hearts card as the first-born twin and the spades card as the second-born twin. For example, if there were 3-year-old twins, and we pulled the 3 of hearts from the deck, that would correspond with the 3-year-old twin who was born first.

After completing the demographics form, RAs provided instructions for the pretest measures and left the room while participants completed them, so as to avoid influencing their responses. The pretest measures assessed participants’ attitudes toward spanking (Attitude Toward Spanking; Holden, Coleman, & Schmidt, 1995), reported use of various discipline strategies over the previous seven days (Parental Responses to Child Misbehavior; Holden, & Zambarano, 1992), perceptions of the extent to which they use various positive parenting practices (Webster-Stratton, 2016), and their perception of the degree to which the target child
demonstrates aggression, defiance, and attention problems (Child Behavior Checklist, abbreviated version; Achenbach & Rescorla, 2000).

**Randomization.** While the pretest measures were being completed, RAs performed the stratified randomization procedure based on participants’ race/ethnicity as reported on the demographic measure. The purpose of stratifying the randomization process was to ensure that the treatment and control groups were equally racially diverse, because an imbalance in demographic characteristics between groups could bias results. In a separate, private workspace reserved for research activities (across the hall from participants), RAs had two randomly generated lists of 0’s and 1’s. The number “0” designated that the person would be assigned to the control group, and the number “1” designated that they would be assigned to the treatment group. The RA looked at the participant’s response to the race/ethnicity item on the demographic measure. For participants who selected “White,” the RA used the randomization list for participants who identify as White; for participants who selected *any race/ethnicity other than* “White,” the RA used the randomization list for participants who identify as any race/ethnicity other than White. The RAs worked down each list; for example, if the first open line was next to a “1” the participant was designated to the treatment group. The RA wrote their participant ID number – designated in numerical order, thus identified by the previous participant’s number (e.g., 026) plus 1 (e.g., 027) – on the line next to the “0” or “1” that determined their condition assignment, and then circled the “0” or “1” at the top of their demographic measure to indicate the condition to which the participant had been assigned. The next participant was assigned based on the next number on the appropriate list.
Intervention

After randomization, the RA used a laptop to prepare the website relevant to the condition to which the participant was assigned. The RA reentered the testing room with the laptop, delivered verbal instructions for the relevant condition, and provided a physical copy of the relevant instructions.

Treatment group. For participants in the treatment group, the RA gave instructions for using the Play Nicely intervention. Play Nicely is a computer-based program designed to help caregivers effectively manage child aggression (Scholer & Goad, 2003). The full program includes approximately 50 minutes of video and audio information on addressing instances of child aggression. However, consistent with previous studies that evaluated Play Nicely’s effects on caregivers’ attitudes toward spanking, the present study used only the interactive section of the program as the intervention. This section of Play Nicely describes a hypothetical scenario in which a child hits another child and the user is shown a menu of 20 options for responding to the situation (e.g., say “no,” time-out, give a warning, take away a privilege; see Figure 2).

Upon clicking on an option, the user is told if they chose “a great option,” “a good option after others have failed,” or if “there are better options,” with an explanation as to why each response is effective or ineffective (see Figure 3). The format of the explanation varies between options, such that most explanations consist of an audio clip and visual text while some include video as well.
For example, the “spank your child” option displays a few bullet points of text (image below) that summarize the following audio clip that plays simultaneously:

“Spanking may seem to help in the short run, but it can actually make matters worse in the long run. Spanking may get a child’s attention, but it does nothing to explain why the behavior is wrong. You are your child’s role model. If you spank them, they will learn from you that physically hurting someone else is an acceptable way to deal with a challenging situation. It does not make sense to teach a child not to hit by hitting them. It is much better to use other ways, such as redirecting, to teach your child how not to hurt people. Because there are much better options to consider, spanking is not recommended as a form of discipline. Some justify spanking by quoting the biblical phrase that says, ‘Spare the rod, you
spoil the child.” However, shepherds usually used a rod to guide sheep—not to hit them. Children should not be guided with being hit or spanked. Rather, they should be guided by setting the rule, redirecting behavior, teaching why hurtful behavior is wrong, and getting them to think about the feelings of others. It is likely that if you spare the guidance, your child will develop behavior problems.”

![Figure 3. Visual information presented in Play Nicely when “Spank your child” option is selected](image)

Participants were asked to select and read/listen to the explanation of all 10 options in the left column of the menu (to ensure that they were exposed to the information about spanking) and any six of the options in the right column of the menu (viewing 16 options in total), which took approximately 20 minutes. For participants 1-76, the RA left the room while the participant completed the intervention procedures. Due to several instances in which RAs returned to find participants on a phone call or otherwise inattentive toward the intervention, this procedure was revised such that an RA sat quietly in the exam room while the intervention was completed (for participants 77-107). Along with this revised procedure, RAs began recording how many of the Play Nicely options each participant actually viewed, as indicated by icons that appear on the main menu.
Control group. Participants in the control group were instructed to view a website about child safety (http://www.safekids.org/safetytips/field_age/little-kids-1–4-years) for 20 minutes. This website includes 21 categories of safety information for kids ages 1-4. Users can click on each of these categories to read tips about various aspects of safety (e.g., choking prevention, fire safety, playground safety, sports safety) for young children. Participants in this group were asked to select and read the information from 16 of the categories on this website, or read as many categories as possible within 20 minutes.

After a participant viewed either the Play Nicely program or the child safety website for 20 minutes, the RA administered the posttest measures and then left the room while they were completed. At this stage, participants completed an adapted version of the Multicultural Therapy Competency Inventory (MTCI) to assess their perception of the intervention’s cultural sensitivity. Participants then completed the Attitude Toward Spanking (ATS) questionnaire again, in order to test for any immediate effects on their attitudes about spanking. Participants were given a $20 gift card for completing the first portion of the study, provided with study personnel contact information, and escorted back to the waiting room.

Follow-up data for all participants was collected one month after recruitment via telephone. Follow-up measures for both conditions included the Parental Responses to Child Misbehavior (PRCM) questionnaire to assess for changes in the reported use of CP and other discipline strategies, then the ATS questionnaire for a third time. All participants who completed the second portion of the study (i.e., follow-up phone call) were sent a second $20 gift card in the mail (i.e., could earn up to $40 in total).
Measures

**Demographics.** A demographic measure was administered at pretest. This measure presented participants with multiple-choice questions regarding their marital status, race, education level, employment status, the education level and employment status of their spouse or live-in partner (if applicable), and their gross annual household income. It also included items for participants to indicate the age and gender of their target child, as well as the ages and genders of anyone else living in their household at least 50% of the time (including the participant), and their relationship to the target child (e.g., brother, aunt).

**Incredible Years Parent Practices Interview** (PPI; Webster-Stratton, 2003). The PPI was developed to evaluate progress in the Incredible Years Program. This 73-item instrument uses Likert scales to measure parents’ self-reported perceptions and use of behaviors that promote desirable behaviors from their children (Webster-Stratton, n.d.). The PPI was designed to be administered to parents of children ages 2-6 years old and was developed to evaluate progress in an evidence-based parenting program called the Incredible Years. This measure is made available for free on the author’s website. The website also provides a scoring guide, which indicates internal reliability scores of 0.75 for the Positive Verbal Discipline scale, 0.67 for the Praise and Incentives scale, and 0.66 for the Clear Expectations scale (Webster-Stratton, 2003). For the purpose of this study, questions on less relevant topics were removed, leaving the 22 items that make up summary scales of Positive Verbal Discipline, Praise and Incentives, and Clear Expectations. All 22 items have a maximum score of 7. Four items have a minimum score of 0 (indicating an “N/A” response) and the remaining 18 items have a minimum score of 1. Six items are reverse scored. Item scores within each scale are averaged to produce component scores ranging between 1-7. Higher scores indicate higher reported use and/or endorsement of
praise, rewards, and clear expectations. This measure was added after 76 participants had
finished the study and thus was only completed by 31 participants; these data are therefore only
presented as descriptive statistics to provide context for the nature of (some of) the sample at
baseline.

**Child Behavior Checklist for Ages 1½-5.** The CBCL is a 100-item measure that asks
caregivers to rate their preschool-aged (i.e., 1.5- to 5-year-old) child’s problem behaviors
(Achenbach & Rescorla, 2000). Response options for each item are 0, 1 and 2, which stand for
“not true (as far as you know),” “somewhat or sometimes true,” and “very true or often true,”
respectively. The items on this measure load onto seven empirically based scales of common
problem areas for children as well as five scales consistent with DSM-5 diagnostic categories,
which can be used to compute a Total Problems score. The CBCL’s internal reliability is strong
for the DSM scales (Cronbach’s $\alpha = .80$), the empirically based scales (Cronbach’s $\alpha = .80$), and
the Total Problems scale (Cronbach’s $\alpha = .95$ to .97). The mean test-retest reliability is high ($r = .83$) for the DSM scales, for the empirically based scales ($r = .85$), and for Total Problems ($r = .91$; Achenbach, Dumenci, & Rescorla, 2003). For the purpose of this study, participants
responded to the 22 items that correspond to two of the empirically based scales that combine to
yield an “Externalizing Problems” composite score (i.e., Attention Problems and Aggressive
Behavior), with a possible score range of 0-44. For the Externalizing Problems composite, raw
scores between 21-24 are in the borderline clinical range, and raw scores of 25 or higher indicate
the presence of clinically significant externalizing problems. Caregivers with higher
scores perceive problems with attention, aggression, and/or defiance from the target child.

Following data collection, factor analyses indicated that 9 of the 22 CBCL items administered
did not contribute substantial explanatory value to the construct model, so subsequent CBCL
analyses included data for the 13 items that fit the model (thus shifting possible score range to 0-26).

**Attitude Toward Spanking survey.** The ATS survey uses a 7-point Likert scale to measure the extent to which participants agree or disagree with 10 statements about spanking (Holden, Coleman, & Schmidt, 1995). Each item can be scored as 1-7 points, where 1 means strong agreement and 7 means strong disagreement. Four items are reverse-scored, then the points for each item are summed to produce a total score with a possible score range of 10-70 (Vittrup et al., 2006). High scores indicate more positive attitudes toward spanking, while low scores indicate less approval of spanking. Parents’ scores on this measure have been found to be significantly correlated with their subsequent spanking behavior (Vittrup et al., 2006). This measure has high internal reliability (ranging from .88 to .90) and test-retest reliability over a 2-week period (r = .76; Holden et al., 1995). Participants completed this measure at pretest, posttest, and follow-up. Factor analyses indicated that 8 of the 10 ATS items meaningfully contributed to this construct model, so analyses of ATS data included those 8 items, resulting in a revised score range of 8-56.

**Parental Responses to Child Misbehavior.** The PRCM lists 9 disciplinary methods (e.g., withdrawal of privileges, spank, use time-out) and asks caregivers to report how many times they used each option in response to misbehavior from the target child during the last week (Holden & Zambarano, 1992). Each discipline strategy is listed as an item with 6 possible responses labeled A-F, where A means none, B means 1-2 times, C means 3-4 times, and so on, up to F meaning 9 or more times. Rather than obtaining a total score, the purpose is to determine how frequently caregivers report using each strategy. This measure has adequate test-retest reliability (r = .64 over a 3-week period; Touliatos, Perlmutter, & Holden, 2001). Caregivers’
reported use of spanking was assessed with item #6 on this measure. Participants completed this measure at pretest and at one-month follow-up.

**Multicultural Therapy Competency Inventory.** The client version of the MTCI is a 32-item instrument used to assess clients’ perceptions of their therapists’ cultural competence. Each item can be answered with “does this very well,” “does this adequately,” or “does this poorly.” Research indicates that the MTCI is reliable and valid, and predicts client satisfaction with treatment (Cole, Piercy, Wolfe, & West, 2014). For the present study, relevant items were adapted to refer to Play Nicely rather than a therapist (e.g., “The training explains things in a way that demonstrates familiarity with my family’s culture.”), resulting in a 15-item adapted measure with a possible score range of 0-30. Higher scores indicate more positive perceptions of an intervention’s cultural sensitivity. Participants completed this measure at posttest. Factor analyses revealed that 7 of the 15 survey items appropriately fit the construct model, so only those 7 items were retained for analyses of MTCI data.

**Results**

**Descriptive Statistics**

**Target child externalizing behaviors.** Of a possible range of 0-44 points on the CBCL “Externalizing Problems” scale (including scores for all 22 survey items here for contextual comparison), mean scores were 13.23 (SD=7.45) for the control group and 16.29 (SD=10.27) for the treatment group. These means are below the threshold of clinical significance based on CBCL scoring criteria, indicating that participants in this sample did not endorse elevated levels of externalizing behaviors for their target child.

**Positive parenting practices.** PPI data collected from participants 77-107 (n=31) were computed to yield component scores for three summary scales, with a possible range of 1-7. The
overall mean of these component scores for the Positive Verbal Discipline scale was 5.28 (SD=1.13); comparison data provided in the PPI scoring guide (n=657) indicates a mean of 5.34 (SD=0.88) for this scale (Webster-Stratton, 2003). The overall mean for the Praise and Incentives scale for the current sample was 4.33 (SD=0.70); the comparison data for this scale indicates a mean of 4.29 (SD=0.82). The sample’s mean score for the Clear Expectations scale was 5.49 (SD=1.50); the comparison data reported for this scale gives a mean of 3.88 (SD=0.91). The present sample’s mean component scores were on the high end of the possible range for all three scales, indicating that participants in this sample reported high levels of use and/or endorsement of praise, rewards, and clear expectations. The current sample’s scores were similar to comparison data for the Positive Verbal Discipline and Praise and Incentives scales, and their score on the Clear Expectations scale was higher than that of the comparison sample.

**Attitudes toward spanking.** With a possible score range of 10-70 points (including data for all 10 items at this point for contextual comparison), the sample’s mean ATS scores at baseline were 32.38 (SD=13.34) for the control group and 29.30 (SD=13.46) for the treatment group. These scores fall into the lower half of the possible range, indicating that participants showed a somewhat low endorsement of favorable attitudes toward spanking. Pretest ATS means reported by Scholer et al. (2010) utilized a different scoring scale (i.e., possible range of 10-50) so a direct comparison cannot be made, but their baseline means were similar to the scores of the current sample.

**Perceptions of the intervention’s cultural sensitivity.** The MTCI mean score for the treatment group was 23.64 (SD=6.11) out of a possible 30 points (including data for all 15 items here). This value cannot be compared with data reported in previous research because the current project employed an adapted version of the MTCI survey with fewer items than the original
instrument, but the group mean falls in the high end of the possible range, indicating that the
treatment group had an overall positive perception of Play Nicely’s cultural appropriateness.

**Compliance with intervention engagement.** For participants 77-107, an RA remained
in the exam room during the intervention. Afterward, for participants in the treatment group, the
RA made record of the number of Play Nicely options the participant had viewed, as indicated
by icons that appear on the main menu. Participants were instructed to engage with 16 of the 20
options. Of the 31 participants for whom this data was recorded, 10 were in the control condition
and therefore were not scored. Of the remaining 21 participants: 19 (90.48%) viewed 16 options,
1 (4.76%) viewed all 20 options, and 1 (4.76%) viewed 6 of the options. This subgroup
represents only 29% of the total sample, but their high degree of engagement is promising.

**Preliminary Analyses**

**Strengthening constructs.** Goodness-of-fit testing and factor analyses were conducted
on the following latent variables of interest: favorable attitudes toward spanking (as measured by
the ATS survey), target child externalizing behaviors (as measured by the CBCL), and
perception of the intervention’s cultural sensitivity (assessed by the MTCI). The purpose of this
was to identify and remove survey items with unacceptably low factor loadings to reduce
statistical noise before creating index scores for participants (Schreiber, Nora, Stage, Barlow, &
King, 2006). Starting with models that included all survey items for each construct, a backward
elimination procedure was employed. Recommendations for factor loading cut-offs vary, but
textbooks by Hair, Anderson, Tatham, and Black (1998) and Stevens (2002) both indicate that a
sample size of 100 subjects necessitates factor loadings of at least ±0.5 in order to have practical
and/or statistical significance (with less stringent requirements for larger samples). The survey
item with the lowest loading score was removed from the model—permanently, if the model was
improved by its removal—and this process was repeated until the model met minimum cut-off values on several goodness-of-fit indicators (see Table 2). There were a few instances in which removing an item with a low factor loading (e.g., ATS item #9=0.473) negatively impacted the overall model, so these items were put back into their respective models. This process resulted in the removal of two items from the ATS scale, nine items from the CBCL, and eight items from the MTCI, and yielded well-fitting models for all three constructs. Note that our measure of reported use of spanking is based on just one item of the PRCM survey, so this procedure was not carried out on that construct. See Figures 4, 5, and 6 for final model and factor loadings of each construct.

**Standardization.** As demonstrated by the factor analyses, some items within the ATS survey, for example, have a stronger relationship with the corresponding latent variable than others. Even though the lowest-loading items were removed, to simply sum a participant’s item scores to create an “ATS Pretest composite” would give equivalent weight to each of the items included, and thus would not appropriately represent the relative contributions of each item score. In light of this, participants’ revised raw scores (after items were removed) were weighted using the standardized regression weight identified for each observed variable and combined to produce a standardized factor score for each participant (for the CBCL and the MTCI, both measures only administered once). Because participants completed the ATS survey on three occasions, it was important to ensure that the same underlying construct was being measured across timepoints (e.g., Item 1 carries the same weight at pretest, posttest, and follow-up) to allow for meaningful comparisons to be made. Thus, participants’ raw ATS scores were weighted as described above and constrained together to make each item’s weight consistent across timepoints, and then combined to produce a standardized factor score for each
participant’s ATS data at pretest, posttest, and follow-up. Additional reasons for standardizing scores for these three constructs were that there is no true scale underlying any of these latent variables, and the scales utilized in this study differ across the constructs, so expressing the data in terms of standard deviations puts the constructs on an equivalent scale. Transforming the data put the overall sample mean at (or close to) 0 and the overall sample’s standard deviation from the mean at (or close to) 1 for each construct. See Table 3 for a summary of ATS raw scores, revised raw scores, and standardized scores. Analyses were conducted on all three versions of ATS data for comparison (see Aim 1 results). All subsequent analyses were conducted using the standardized scores.

**Outlying and missing data.** All 107 participants completed the treatment to which they were randomly assigned. Three ATS change scores were identified as outliers, but parameter estimates with and without their inclusion were equivalent, so these observations were included in the reported analyses. Seven participants were unable to be reached for follow-up data collection. In order to assess whether non-completion of data collection was associated with a relevant variable, 2x2 ANOVAs were conducted using ATS pretest data as the dependent variable. There was no significant interaction between “completion” and Condition ($F(1, 103)=0.200, p=0.655$) or between completion and Race ($F(1, 103)=1.884, p=0.173$). In the absence of any other known variable that could be associated with their missingness, these data are assumed to be missing completely at random (MCAR). Additionally, no difference was found between main analyses that excluded participants who did not provide follow-up data versus intention-to-treat (ITT) analyses, so only ITT results are reported (i.e., analyses include all data from all participants, with group assignment designated by original randomization procedure).
Baseline comparisons. Chi-square analyses and independent samples t-tests were used to assess for baseline differences between the treatment and control groups’ demographic characteristics and scores on pretest measures. Results indicated that there were no statistically significant differences in participants’ demographic variables (i.e., race, marital status, gender, relationship to target child, education level, gross annual household income) between conditions. Regarding pretest measures, there was not a significant difference between conditions’ reported use of spanking in the last week (as measured by item #6 on the PRCM scale, see Table 4; \( \chi^2(3) = 4.999, p = 0.172 \)). Baseline ATS scores were not statistically different between conditions \((t(105) = -1.527, p = 0.130)\) or between White participants and participants of color within the treatment group \((t(47) = -0.042, p = 0.967)\). There was a statistically significant difference between the groups’ CBCL scores at baseline \((t(240.62) = 3.207, p = 0.002)\), with higher scores in the treatment group indicating treatment participants reported higher rates of externalizing behavior among their children than did the control group. Out of concern that this variable could influence outcomes regarding attitudes toward and/or reported use of spanking, all main analyses were conducted both with and without CBCL data as a control variable; CBCL data did not influence outcomes for any of the models so this variable was excluded from the results reported below.

Primary Outcomes

Aim 1: Compare changes in attitudes toward spanking between conditions. A linear regression model with mixed effects was used to examine participants’ scores on the ATS survey (a continuous dependent variable), measured on three occasions (i.e., pretest, posttest, follow-up). Fixed effects included Time, Condition, and the interaction term Time by Condition. Time alone did not significantly contribute to the model \((F(2, 207.36) = 0.012, p = 0.988)\). Condition did have a significant contribution \((F(1, 107.12) = 5.501, p = 0.021)\), as did the Time by Condition
interaction \((F(2, 207.36)= 3.188, p=0.043)\). Thus, including these predictors made the regression model fit the data better than the model with no independent variables included. Estimates of fixed effects parameters revealed that changes in ATS scores were not significantly different between conditions from pretest to posttest \((t(207.10)= -1.820, p=0.070)\) or from posttest to follow-up \((t(207.50)= -0.642, p=0.521)\). Results did show, however, that changes in ATS scores were significantly different between conditions from pretest to follow-up \((t(207.50)= -2.415, p=0.017)\). The “difference in the differences” was -0.299 standard deviations (a negative value because a lower ATS score indicates less favorable attitudes toward spanking), 95% CI [-0.544 to -0.055], indicating a small to medium effect.

The same analysis was run using the sample’s original raw ATS scores (including all 10 survey items). Results from this analysis found significant contributions for all three fixed effects: Time \((F(2, 207.64)= 18.059, p=0.000)\), Condition \((F(1, 107.20)=5.914, p=0.017)\), and Time by Condition \((F(2, 207.64)= 4.209, p=0.016)\). Estimates of fixed effects showed a significant difference in the changes in ATS scores between conditions from pretest to posttest \((t(207.16)= -2.519, p=0.013)\) and from pretest to follow-up \((t(207.89)= -2.484, p=0.014)\), but not from posttest to follow-up \((t(207.89)= -0.029, p=0.977)\). The analysis was also run on the sample’s revised raw ATS scores (after removing two survey items, but before weighting and standardizing scores). Results from this analysis similarly found significant contributions for the fixed effects of Time \((F(2, 207.68)= 19.695, p=0.000)\), Condition \((F(1, 107.23)= 5.825, p=0.017)\), and the Time by Condition interaction \((F(2, 207.68)= 4.756, p=0.010)\). The estimates of fixed effects parameters using revised raw scores also indicated a significant difference between conditions’ changes in ATS scores from pretest to posttest \((t(207.20)= -2.677, p=0.008)\), and from pretest to follow-up \((t(207.94)= -2.641, p=0.009)\), but no significant difference from
posttest to follow-up ($t(207.94)=-0.032, p=0.974$). These outcomes may indicate that the treatment group did demonstrate a change in ATS scores immediately at posttest, though this was not detected when standardized scores were analyzed. They also indicate that revisions to the data (i.e., revising constructs and standardization) maintained the data’s integrity, as outcomes were comparable (though differing in magnitude) across these analyses.

**Aim 2: Compare changes in reported use of spanking between conditions.**

Participants’ reported use of spanking was measured by item #6 on the PRCM scale (henceforth referred to as "PRCM6"). The sample’s responses to this item at follow-up created logistical challenges, as every participant in the treatment group selected either “0” or “1-2 times” (regarding number of times they spanked their child in the last week). Because of this, PRCM6 data for the whole sample had to be dichotomously recoded to either “0” or “1 or more.” This now binary outcome variable could not fit a normal distribution, so it was analyzed using a mixed effects logistic regression with random intercepts by subject. This test produces odds ratios which, in this case, indicated the odds of having a score of 1 (i.e., reported spanking child 1 or more times in the last week) in the treatment group versus the control group. Fixed effects again included Time, Condition, and an interaction of Time by Condition. Tests of fixed effects indicated that none of these terms added predictive value to the model. No significant main effects were detected for Time ($F(1, 203)=1.715, p=0.192$) or Condition ($F(1, 106)=1.167, p=0.282$), and the Time by Condition interaction was insignificant as well ($F(1, 203)=0.002, p=0.965$).

**Aim 3A: Compare changes in attitudes toward spanking between POC and White participants in the treatment group.** This analysis aimed to essentially repeat the Aim 1 analysis for participants in the treatment condition only, using Race as the grouping variable (i.e.,
“White” or “POC” rather than “treatment” or control”) to assess whether treatment group participants’ changes in ATS scores varied by racial identity. Fixed effects included Time, Race, and Time by Race interaction. No significant contributions were detected for Time ($F(2, 94.221) = 1.442, p = 0.242$), Race ($F(1, 49.019) = 1.435, p = 0.237$), or the Time by Race interaction ($F(2, 94.221) = 0.001, p = 0.999$), thus indicating that changes in attitudes toward spanking (and/or lack thereof) among the treatment group were consistent between White participants and POC.

**Aim 3B: Compare changes in reported use of spanking between POC and White participants in the treatment group.** Like the above, this subgroup analysis of the treatment condition aimed to repeat the Aim 2 procedure. Race was used as the grouping variable (rather than treatment or control) and reported use of spanking (recoded to 1 = 1 or more times last week, 0 = 0 times last week) was the outcome variable, for the purpose of assessing whether treatment group participants’ outcomes varied by race. Fixed effects included Time, Race, and a Time by Race interaction term. The main effect of Time was not significant ($F(1, 90) = 0.272, p = 0.603$), nor was the Time by Race interaction term ($F(1, 90) = 0.380, p = 0.539$). There was, however, a significant main effect detected for Race ($F(1, 70) = 4.696, p = 0.034$), indicating that this term may be a predictor of reported use of spanking. Fixed-effects regression coefficients showed that the odds of reportedly spanking in the last week were higher for participants of color (across timepoints), but the odds ratio was not significantly different from 1:1 ($B = 0.912$, SE = 0.78, $p = 0.25$) so a causal statement cannot be made about the effects of each level of Race on likelihood to endorse use of spanking.

**Aim 4: Compare perceptions of the intervention’s cultural sensitivity between POC and White participants in the treatment group.** Within the treatment group, Shapiro-Wilk
tests detected non-normal distributions of MTCI data for both White participants \((p=0.001)\) and POC \((p=0.003)\), indicating that a parametric analysis would not be appropriate. Visual inspection of a population pyramid chart showed that the subgroups’ MTCI scores had similarly shaped distributions. Therefore, a nonparametric Mann-Whitney \(U\) test was used to assess for a difference in perceptions of the intervention’s cultural sensitivity between White participants and POC. The median MTCI scores for participants of color \((Mdn=0.221)\) and White participants \((Mdn=0.619)\) were not statistically significantly different \((U=309.00, z=0.184, p=0.854)\). This indicates that there was not a substantial discrepancy in perceptions of the cultural sensitivity of Play Nicely between participants of color and White participants in the treatment group.

**Discussion**

A growing research body has sought to address the use of spanking as a discipline strategy, with several studies providing preliminary support for the utility of the brief interactive Play Nicely program in decreasing parents’ favorable attitudes toward spanking. Previous findings were promising enough to warrant further investigation into this application of the program, as well as the evaluation of Play Nicely as a possible intervention for decreasing the use of spanking and examining whether outcomes vary based on racial group. The intention of the current study was to use a randomized controlled trial design to: 1) Build upon extant studies’ evaluation of the effects of Play Nicely on caregivers’ attitudes toward spanking, with a focus on low-income families with young children (1-5 years old); 2) examine whether Play Nicely produces changes in caregivers’ reported use of spanking; 3) assess for differences in the changes in attitudes toward spanking and reported use of spanking between White participants and participants of color within the treatment group; 4) compare perceptions of the cultural sensitivity of the Play Nicely intervention between White participants and participants of color in
the treatment condition. In the service of these goals, a low-income and racially diverse sample of 107 caregivers were recruited from a pediatric clinic to participate in this study.

Results partially support the author’s a priori hypotheses, with mixed results on several domains. First, with regard to attitudes toward spanking and use of spanking, participants’ attitudes toward spanking were significantly less favorable in the treatment group than the control group at follow-up, but not at post-test. Moreover, there was no significant change in participants’ reported use of spanking at follow-up. The current study met its goal of replicating previous evaluations of Play Nicely’s impact on attitudes toward spanking. While the results are not directly comparable, this study detected a small but significant treatment effect, using a sample of similar size to that of Scholer et al. (2010) and less than half the size of Chavis et al. (2013). The study employed and maintained relatively stringent methodological procedures that allowed for comparison of ATS scores over three timepoints. Given the substantial efforts that were directed toward maximizing the treatment effect (e.g., more substantial exposure to the intervention) the confirmation of past findings that Play Nicely reduces favorable attitudes toward spanking is unsurprising. We hypothesized, though, that this change in attitudes would be accompanied by a change in reported use of spanking. It is possible that the lack of significant effects on spanking behavior could be explained by the small magnitude of the difference between groups’ changes in attitudes toward spanking, such that perhaps a change in behavior would only follow attitude changes of a greater magnitude. Additionally, reported use of spanking at pretest indicated that spanking was a relatively low-base rate behavior for this sample. Detecting a decrease in an already low-base rate behavior across groups requires enormous statistical power, so it is possible that an existing effect was unable to be detected. It
also could be the case that a change in behavior was not detected because of measurement error, as this outcome was based on participants’ self-reported recall of spanking over the last week.

In terms of the comparative outcomes between White participants and POC, there were no differences in the groups’ perceptions of Play Nicely’s cultural sensitivity or changes in attitudes toward spanking—the latter of which diverges from the findings of the pilot for this study conducted by the current author, wherein White participants demonstrated a substantially greater decrease in ATS scores after intervention (Richardson & Damashek, 2019). While this is a desirable outcome that may simply indicate that Play Nicely is well-received and equally effective across racial groups, it is important to consider other possible explanations—especially given that the sample of that pilot was a subset of the present study’s sample. Recall that some literature posits a “stress theory,” which indicates that higher rates of spanking are associated with single parenting and low income (for example) as well as being African American because of the increased stress often experienced by individuals in these groups (Giles-Sims et al., 1995). The sample of the current study was slightly less White (5.3%) and had slightly higher rates of low income and low education level than did the pilot sample, so these factors may have contributed to participants’ attitudes and offset the difference in scores between racial groups that was previously detected. On the other hand, neither the White participants nor the POC subgroup reported a significant change in their use of spanking, but Race was found to be a predictor of this outcome. This finding may indicate that, while racial identity can impact the likelihood of reportedly using spanking, there were no differences between the subgroups’ changes in attitudes/behavior because neither subgroup demonstrated very substantial a change. Additional research on the outcomes of Play Nicely comparing race/ethnicity as well as income
level and other demographic characteristics is needed to clarify the presence (or absence) of the effectiveness of this intervention across diverse backgrounds.

**Limitations**

Despite this study’s strengths it faced numerous challenges, many of which were related to an attempt to measure participants’ “use of spanking.” As has been discussed in the literature, the practical limitations around obtaining accurate, objective data on caregivers’ use of spanking remains a substantial barrier to progress in this area of study. The instruments selected to measure outcome variables were among few of their kind. It is unlikely that any pen-and-paper instrument could yield an accurate measure of parents’ use of spanking, due to the aforementioned issue of self-report bias, but the PRCM6 presented additional unforeseen issues. The measure’s response categories (e.g., “1-2 times,” “9 or more times”) precluded the data from being readily analyzable; specificity of data is lost in these categories, and this sample’s lack of responses in particular categories required data to be even further diluted into a dichotomous comparison. Another limitation was that our randomized stratification procedure allocated more participants to the control group than the treatment group, further limiting the statistical power of analyses within the treatment condition. Additionally, the size of the sample necessitated compiling participants who endorsed any racial/ethnic identity other than “White” into one subgroup for analyses (i.e., the “participants of color” subgroup), rather than examining outcomes across the distinct racial/ethnic identities represented in the sample. Finally, this sample was from one community in the Northern U.S. that is relatively progressive (which might explain the low reported base rate of spanking), so a lack of geographic diversity may limit the generalizability of these findings.
Implications

The study of spanking is complex and requires innovative methodologies for both measurement and intervention. Research on the Play Nicely program has consistently shown that its use results in caregivers viewing spanking less favorably, and it is conceptually an attractive option as a spanking intervention due to its brevity and low-barrier access. There remains, however, a lack of empirical evidence for this or any brief, accessible intervention’s ability to change caregivers’ actual use of spanking as a discipline strategy. Furthermore, recent studies have corroborated existing evidence that racial/ethnic identity and other sociodemographic factors have a meaningful impact on caregivers’ attitudes and behaviors related to spanking, and thus these factors must be a focus of intervention and research development in this area.

Researchers examining spanking in future studies are encouraged to consider creative uses of technology as alternative solutions for collecting objective data on spanking behavior (e.g., texting participants once a week to ask how many times they have spanked *that day*). It may also be useful to collect qualitative data in this area, surveying would-be users (i.e., caregivers) for qualities they would like or dislike about a relevant intervention, what kind of information (or vehicle thereof) might impact their openness to changing their parenting approach, et cetera. Finally, additional research on Play Nicely and other potential approaches to spanking reduction must examine the extent to which the intervention serves families across race/ethnicity and other cultural factors. Moreover, this line of study would benefit from attempts to identify compensatory solutions that ameliorate outcome discrepancies, thus promoting the development of equitable and accessible parenting interventions for a broad range of families.
References


Family Relations, 56, 80-91.


http://www.playnicely.org/.


<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>Control group (n=58)</th>
<th>Treatment group (n=49)</th>
<th>Total sample (n=107)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age of participants in years</td>
<td>31.31</td>
<td>30.70</td>
<td>31.03</td>
</tr>
<tr>
<td>Gender of participant (female)</td>
<td>55 (94.8%)</td>
<td>44 (89.8%)</td>
<td>99 (92.5%)</td>
</tr>
<tr>
<td>Mean age of target child in months</td>
<td>38.70</td>
<td>35.87</td>
<td>37.42</td>
</tr>
<tr>
<td>Gender of target child (male)</td>
<td>33 (56.9%)</td>
<td>24 (49.0%)</td>
<td>57 (53.3%)</td>
</tr>
<tr>
<td>Mean # of children in home</td>
<td>2.67</td>
<td>2.43</td>
<td>2.56</td>
</tr>
<tr>
<td>Participant race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>30 (51.7%)</td>
<td>24 (49.0%)</td>
<td>54 (50.5%)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>15 (25.9%)</td>
<td>17 (34.7%)</td>
<td>32 (29.9%)</td>
</tr>
<tr>
<td>Biracial</td>
<td>7 (12.1%)</td>
<td>3 (6.1%)</td>
<td>10 (9.3%)</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>2 (3.4%)</td>
<td>1 (2.0%)</td>
<td>3 (2.8%)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1 (1.7%)</td>
<td>1 (2.0%)</td>
<td>2 (1.9%)</td>
</tr>
<tr>
<td>Asian American</td>
<td>1 (1.7%)</td>
<td>0 (0%)</td>
<td>1 (0.9%)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (3.4%)</td>
<td>3 (6.1%)</td>
<td>5 (4.7%)</td>
</tr>
<tr>
<td>Relationship to target child</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother figure</td>
<td>50 (86.2%)</td>
<td>41 (83.7%)</td>
<td>91 (85.0%)</td>
</tr>
<tr>
<td>Father figure</td>
<td>3 (5.2%)</td>
<td>5 (10.2%)</td>
<td>8 (7.5%)</td>
</tr>
<tr>
<td>Grandparent</td>
<td>2 (3.4%)</td>
<td>2 (4.1%)</td>
<td>4 (3.7%)</td>
</tr>
<tr>
<td>Sibling</td>
<td>1 (1.7%)</td>
<td>0 (0%)</td>
<td>1 (0.9%)</td>
</tr>
<tr>
<td>Education completed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>9 (15.5%)</td>
<td>5 (10.2%)</td>
<td>14 (13.1%)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>15 (25.9%)</td>
<td>12 (24.5%)</td>
<td>27 (25.2%)</td>
</tr>
<tr>
<td>Some college</td>
<td>25 (43.1%)</td>
<td>19 (38.8%)</td>
<td>44 (41.1%)</td>
</tr>
<tr>
<td>College graduate</td>
<td>7 (12.1%)</td>
<td>8 (16.3%)</td>
<td>15 (14.0%)</td>
</tr>
<tr>
<td>Post-undergraduate education</td>
<td>2 (3.4%)</td>
<td>4 (8.2%)</td>
<td>6 (5.6%)</td>
</tr>
<tr>
<td>Gross annual household income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than $5,000 - $9,999</td>
<td>13 (22.4%)</td>
<td>11 (22.5%)</td>
<td>24 (22.5%)</td>
</tr>
<tr>
<td>$10,000 - $19,999</td>
<td>14 (24.1%)</td>
<td>11 (22.5%)</td>
<td>25 (23.4%)</td>
</tr>
<tr>
<td>$20,000 - $29,999</td>
<td>11 (18.9%)</td>
<td>4 (8.1%)</td>
<td>15 (14.0%)</td>
</tr>
<tr>
<td>$30,000 - $39,999</td>
<td>4 (6.8%)</td>
<td>5 (10.2%)</td>
<td>9 (8.4%)</td>
</tr>
<tr>
<td>$40,000 - $49,999</td>
<td>6 (10.4%)</td>
<td>3 (6.1%)</td>
<td>9 (8.4%)</td>
</tr>
<tr>
<td>$55,000 and higher</td>
<td>7 (12.1%)</td>
<td>11 (22.5%)</td>
<td>18 (16.8%)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently married</td>
<td>23 (39.7%)</td>
<td>23 (46.9%)</td>
<td>46 (43.0%)</td>
</tr>
<tr>
<td>Never married</td>
<td>22 (37.9%)</td>
<td>20 (40.8%)</td>
<td>42 (39.3%)</td>
</tr>
<tr>
<td>Living with partner</td>
<td>7 (12.1%)</td>
<td>3 (6.1%)</td>
<td>10 (9.3%)</td>
</tr>
<tr>
<td>Separated</td>
<td>4 (6.9%)</td>
<td>3 (6.1%)</td>
<td>7 (6.5%)</td>
</tr>
<tr>
<td>Divorced or annulled</td>
<td>1 (1.7%)</td>
<td>0 (0%)</td>
<td>1 (0.9%)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (1.7%)</td>
<td>0 (0%)</td>
<td>1 (0.9%)</td>
</tr>
</tbody>
</table>

*Response not provided by 2 participants (1.9%).  †Response not provided by 3 participants (2.8%).
 Response not provided by 1 participant (0.9%).  ‡Response not provided by 7 participants (6.5%).
Figure 4. Path diagram of the final factor analysis model for the Attitudes toward spanking construct (as measured by the ATS survey). Items 3 & 7 removed.

Note: Squares designate observed variables (i.e., response scores for items on the ATS survey). The circle represents the “common factor” or latent variable/construct that cannot be directly observed (i.e., participants’ attitudes about spanking). The arrowed lines pointing from the latent variable indicate the “factor loading” of each survey item (i.e., the correlation between the latent variable and each of the observed variables, with a range of ±0-1). Some of the values here are negative because the corresponding survey items are reverse scored. Values on the right are the “unique factors”—error associated with each observed variable. Curved arrowed lines indicate common variance.
Figure 5. Path diagram of the final factor analysis model for the Target child externalizing behaviors construct (as measured by the CBCL survey). Items 1, 3, 4, 11, 13, 15, 17, 20, & 21 removed. 
(See Figure 4 for diagram interpretation information.)
Figure 6. Path diagram of the final factor analysis model for the Perception of intervention’s cultural sensitivity construct (as measured by the MTCI survey). Items 4, 6, 8, 9, 10, 12, 13, & 14 removed.

(See Figure 4 for diagram interpretation information.)
<table>
<thead>
<tr>
<th>Model fit index</th>
<th>General rule for acceptable fit</th>
<th>Construct result</th>
</tr>
</thead>
</table>
| Chi-square test of model fit ($\chi^2$) | Ratio of $\chi^2$ to $df \leq 2$ | ATS = 1.314     
                                                                 CBCL = 1.234     
                                                                 MTCI = 1.149      |
| Tucker-Lewis index (TLI) | $\geq 0.95$ for acceptance | ATS = 0.976     
                                                                 CBCL = 0.964     
                                                                 MTCI = 0.988      |
| Comparative fit index (CFI) | $\geq 0.95$ for acceptance | ATS = 0.985     
                                                                 CBCL = 0.970     
                                                                 MTCI = 0.992      |
| Root mean square error of approximation (RMSEA) | < 0.06 | ATS = 0.054     
                                                                 CBCL = 0.047     
                                                                 MTCI = 0.037      |

* Attitudes toward spanking (as measured by the ATS survey).  * Target child externalizing behaviors construct (as measured by the CBCL).  * Perception of intervention’s cultural sensitivity construct (as measured by the MTCI).  * Recommended goodness-of-fit indicators and corresponding cutoff scores drawn from Schreiber et al. (2006).
### Table 3. ATS data by group and subgroup means (and SDs) across timepoints, in terms of Raw scores, Revised raw scores, and Standardized scores.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Raw scores</th>
<th></th>
<th>Follow-up (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td></td>
</tr>
<tr>
<td><strong>Total sample (n=107)</strong></td>
<td>30.97 (13.42)</td>
<td>29.28 (13.41)</td>
<td>26.41 (13.03)</td>
</tr>
<tr>
<td><strong>Treatment group (n=49)</strong></td>
<td>29.30 (13.46)</td>
<td>25.52 (11.99)</td>
<td>22.36 (9.89)</td>
</tr>
<tr>
<td>White participants (n=24)</td>
<td>27.13 (14.85)</td>
<td>22.50 (13.10)</td>
<td>21.39 (11.29)</td>
</tr>
<tr>
<td>Participants of color (n=25)</td>
<td>31.38 (11.91)</td>
<td>28.42 (10.27)</td>
<td>23.36 (8.33)</td>
</tr>
<tr>
<td><strong>Control group (n=58)</strong></td>
<td>32.38 (13.34)</td>
<td>32.47 (13.81)</td>
<td>29.73 (14.38)</td>
</tr>
<tr>
<td>White participants (n=30)</td>
<td>30.81 (14.07)</td>
<td>31.30 (14.83)</td>
<td>29.96 (14.90)</td>
</tr>
<tr>
<td>Participants of color (n=28)</td>
<td>34.07 (12.55)</td>
<td>33.71 (12.79)</td>
<td>29.48 (14.09)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Groups</th>
<th>Revised raw scores</th>
<th></th>
<th>Follow-up (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td></td>
</tr>
<tr>
<td><strong>Total sample (n=107)</strong></td>
<td>26.06 (11.70)</td>
<td>23.92 (11.63)</td>
<td>21.90 (11.19)</td>
</tr>
<tr>
<td><strong>Treatment group (n=49)</strong></td>
<td>24.73 (11.83)</td>
<td>20.64 (10.00)</td>
<td>18.38 (8.38)</td>
</tr>
<tr>
<td>White participants (n=24)</td>
<td>22.96 (12.72)</td>
<td>18.54 (11.11)</td>
<td>17.17 (9.13)</td>
</tr>
<tr>
<td>Participants of color (n=25)</td>
<td>26.42 (10.89)</td>
<td>22.66 (8.55)</td>
<td>19.64 (7.52)</td>
</tr>
<tr>
<td><strong>Control group (n=58)</strong></td>
<td>27.19 (11.57)</td>
<td>26.69 (12.25)</td>
<td>24.78 (12.39)</td>
</tr>
<tr>
<td>White participants (n=30)</td>
<td>26.10 (12.10)</td>
<td>25.73 (13.06)</td>
<td>25.39 (12.90)</td>
</tr>
<tr>
<td>Participants of color (n=28)</td>
<td>28.36 (11.07)</td>
<td>27.71 (11.46)</td>
<td>24.15 (12.04)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Groups</th>
<th>Standardized scores</th>
<th></th>
<th>Follow-up (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td></td>
</tr>
<tr>
<td><strong>Total sample (n=107)</strong></td>
<td>0.00 (1.42)</td>
<td>0.00 (1.43)</td>
<td>0.00 (1.32)</td>
</tr>
<tr>
<td><strong>Treatment group (n=49)</strong></td>
<td>-0.23 (1.33)</td>
<td>-0.35 (1.24)</td>
<td>-0.41 (0.96)</td>
</tr>
<tr>
<td>White participants (n=24)</td>
<td>-0.42 (1.49)</td>
<td>-0.54 (1.41)</td>
<td>-0.53 (1.08)</td>
</tr>
<tr>
<td>Participants of color (n=25)</td>
<td>-0.04 (1.17)</td>
<td>-0.16 (1.04)</td>
<td>-0.29 (0.83)</td>
</tr>
<tr>
<td><strong>Control group (n=58)</strong></td>
<td>0.19 (1.47)</td>
<td>0.29 (1.52)</td>
<td>0.34 (1.47)</td>
</tr>
<tr>
<td>White participants (n=30)</td>
<td>0.13 (1.47)</td>
<td>0.26 (1.57)</td>
<td>0.43 (1.48)</td>
</tr>
<tr>
<td>Participants of color (n=28)</td>
<td>0.25 (1.49)</td>
<td>0.33 (1.50)</td>
<td>0.24 (1.48)</td>
</tr>
</tbody>
</table>

---

*a Attitudes toward spanking (as measured by the ATS survey). b Calculated by summing each participant’s ATS item scores as originally administered (10 items, 7-point Likert scale, possible score range of 10-70) and taking the group average score. c Calculated by summing each participant’s ATS item scores after Items 3 & 7 were removed (8 items, 7-point Likert scale, possible score range of 8-56) and taking the group average score. d Calculated by weighting ATS items based on factor loadings, constraining weights across timepoints, converting to standardized values, and taking the group average score.*
Table 4. PRCM6a responses in frequency (and %) by condition at pretest and follow-up

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Pretest (n=49)</th>
<th>Follow-up (n=45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 times</td>
<td>36 (73.5)</td>
<td>36 (80.0)</td>
</tr>
<tr>
<td>1-2 times</td>
<td>11 (22.4)</td>
<td>9 (20.0)</td>
</tr>
<tr>
<td>3-4 times</td>
<td>1 (2.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>5-6 times</td>
<td>1 (2.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Control group</td>
<td>Pretest (n=58)</td>
<td>Follow-up (n=55)</td>
</tr>
<tr>
<td>0 times</td>
<td>38 (65.5)</td>
<td>40 (72.7)</td>
</tr>
<tr>
<td>1-2 times</td>
<td>13 (22.4)</td>
<td>9 (16.4)</td>
</tr>
<tr>
<td>3-4 times</td>
<td>7 (12.1)</td>
<td>4 (7.3)</td>
</tr>
<tr>
<td>5-6 times</td>
<td>0 (0.0)</td>
<td>2 (3.6)</td>
</tr>
</tbody>
</table>

*a Parental Responses to Child Misbehavior survey item #6, which asks “How many times did you spank your child in the last week?” Response options also included “7-8 times” and “9 or more times” but were not endorsed by any participants.*
APPENDIX

IRB Approval Letter
Date:  April 7, 2016

To:  Amy Damashek, Principal Investigator
     Hilary Richardson, Student Investigator for dissertation

From:  Amy Naugle, Ph.D., Chair

Re:  HSIRB Project Number 15-12-11

This letter will confirm that your research project titled “Examination of an Internet Program to Improve Discipline and Safety Strategies for Parents” has been approved under the full category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., you must request a post approval change to enroll subjects beyond the number stated in your application under “Number of subjects you want to complete the study”). Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination:  December 15, 2016