

## EMPIRICAL ARTICLE

# Treating children's aggressive behavior problems using cognitive behavior therapy with virtual reality: A multicenter randomized controlled trial

Sophie C. Alsem<sup>1,2</sup>  | Anouk van Dijk<sup>2</sup>  | Esmée E. Verhulp<sup>1</sup>  |  
Tycho J. Dekkers<sup>3,4,5,6,7</sup>  | Bram O. De Castro<sup>2</sup> 

<sup>1</sup>Department of Developmental Psychology, Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup>Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

<sup>4</sup>Level, Academic Center for Child and Adolescent Psychiatry, Amsterdam, The Netherlands

<sup>5</sup>Department of Child and Adolescent Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>6</sup>Accare Child Study Center, Groningen, The Netherlands

<sup>7</sup>Department of Child and Adolescent Psychiatry, Amsterdam University Medical Center (AUMC), Amsterdam, The Netherlands

## Correspondence

Sophie C. Alsem, Research Institute of Child Development and Education, University of Amsterdam, PO 15776, 1001 NG Amsterdam, The Netherlands.  
Email: [s.c.alsem@uva.nl](mailto:s.c.alsem@uva.nl)

## Funding information

Netherlands Organization for Scientific Research, Grant/Award Number: 453-15-004/511

## Abstract

This multicenter randomized controlled trial investigated whether interactive virtual reality enhanced effectiveness of Cognitive Behavioral Therapy (CBT) to reduce children's aggressive behavior problems. Boys with aggressive behavior problems ( $N=115$ ;  $M_{\text{age}}=10.58$ ,  $SD=1.48$ ; 95.7% born in Netherlands) were randomized into three groups: CBT with virtual reality, CBT with roleplays, or care-as-usual. Bayesian analyses showed that CBT with virtual reality more likely reduced aggressive behavior compared to care-as-usual for six of seven outcomes ( $ds$  0.19–0.95), and compared to CBT with roleplays for four outcomes ( $ds$  0.14–0.68). Moreover, compared to roleplays, virtual reality more likely enhanced children's emotional engagement, practice immersion, and treatment appreciation. Thus, virtual reality may be a promising tool to enhance CBT effectiveness for children with aggressive behavior problems.

Aggressive behavior problems are the most common form of malfunctioning in school-aged children (Costello et al., 2003). These problems predict adverse outcomes for children later in life (Burkey et al., 2018; Loeber & Farrington, 2000) and have a continuing negative impact on children's environment (McConaughy & Skiba, 1993; Wilson & Lipsey, 2006). Many intervention programs therefore target aggressive behavior problems as they arise in childhood (Lochman & Matthis, 2017). Cognitive behavior therapy (CBT) can reduce aggressive behavior in children (Weisz & Kazdin, 2017), but

intervention effects tend to be modest and heterogeneous (McCart et al., 2006). Effects can be stronger when interventions focus more on exposure to anger and on solving real-life social problems (de Mooij et al., 2020; Landenberger & Lipsey, 2005). Hence, intervention methods that promote ecologically valid practice may enhance effectiveness (Weisz et al., 2019). Interactive virtual reality may be a promising tool to attain this goal. In interactive virtual reality, children can walk around freely, talk to virtual peers, and play games, offering a realistic and engaging environment to practice new skills

**Abbreviations:** CBCL, Child Behavior Checklist; CBT, Cognitive Behavioral Therapy; DSM, Diagnostic and Statistical Manual of Mental Disorders; IRPA, Instrument for Reactive and Proactive Aggression; ISCED, International Standard Classification of Education; TRF, Teacher Report Form.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Child Development* published by Wiley Periodicals LLC on behalf of Society for Research in Child Development.

during therapy (Lindner, 2021). Our feasibility study showed that using virtual reality in CBT was feasible and acceptable for children in routine care, and had the potential to reduce aggressive behavior (Alsem et al., 2021). The aim of the current randomized controlled trial is to investigate whether virtual reality actually *enhances* effectiveness compared to CBT without virtual reality and care-as-usual.

Virtual reality may have three important benefits for CBT with children. First, practicing in virtual reality can enhance children's emotional engagement and immersion, which is important because CBT practice has been found to be most effective when cognitions and skills are practiced in emotionally engaging situations (Suveg et al., 2007). Children should thus ideally practice whilst experiencing feelings of anger (Sukhodolsky et al., 2016). Virtual reality can simulate anger-provoking situations that children encounter in daily life and has been shown to successfully elicit children's anger (Geraets et al., 2021; Verhoef, van Dijk, et al., 2021). It may be more immersive and engaging than roleplay exercises currently used in CBT, as children do not have to rely on their memory or imagination (Park et al., 2011). Supporting this idea, research found that a virtual reality assessment of aggressive behavior better predicted children's real-life aggressive behavior than an imagery-based assessment using hypothetical stories (Verhoef, Verhulp, et al., 2021).

Second, virtual reality can enhance children's treatment appreciation and their perception of the treatment's efficacy. Children with aggressive behavior problems are often not motivated, or even resistant, to treatment (Frick, 2012; Lochman et al., 2019). It is important to enhance these children's treatment appreciation, which has been related to increases in treatment effectiveness (Lochman, Kassing, & Sallee, 2017). As many children grow up surrounded by digital devices, using technology in interventions may have particular appeal and utility to them (Bakker et al., 2016; Weisz et al., 2019). Indeed, using technology (e.g., adding an internet component) in a treatment for children with aggression problems effectively increased children's treatment participation and perceived efficacy (Lochman, Boxmeyer, et al., 2017). Accordingly, our feasibility study showed that children with aggressive behavior problems highly appreciated CBT with virtual reality (Alsem et al., 2021).

Third, virtual reality allows for individually tailored exercises in CBT. Most current CBTs for children with aggressive behavior problems are provided in groups (Lochman et al., 2019). Although group treatments provide a natural context to practice in roleplays with actual peers, they limit opportunities to adjust the exercises to each child's specific needs. Moreover, individual therapy can lead to larger decreases in children's aggression than group therapy (Lochman et al., 2015; Wilson & Lipsey, 2007), whereas group therapy may yield iatrogenic effects (Dodge et al., 2006). Virtual reality provides an opportunity to combine individual therapy with ecologically valid practice

with virtual peers. Focusing the exercises on the situations, cognitions, and behaviors of an individual child can not only enhance children's treatment appreciation and adherence, but also the effectiveness of the intervention (Hollis et al., 2017; Lochman, Kassing, & Sallee, 2017).

Although virtual reality has the potential to enhance effectiveness compared to current CBT treatments using roleplays, no study so far has investigated this (Hadley et al., 2019). Studies with a multi-armed design are needed to investigate the added benefits of virtual reality compared to identical intervention without virtual reality and care as usual (Lindner, 2021). We took this into account by comparing CBT with interactive virtual reality not only to care-as-usual, but also to the same CBT using similarly structured roleplay exercises. We conducted our study within routine care, as the use of virtual reality in therapy is increasingly called for by clinicians (Lindner et al., 2019).

We developed the new individual CBT 'YourSkills' based on evidence-based treatments for children with aggressive behavior problems. YourSkills targets deficits in emotion regulation and social information processing—two mechanisms underlying childhood aggression (Crick & Dodge, 1994; Lochman & Matthys, 2017). Similar to most CBTs for aggression, children learn to monitor their anger and practice techniques to modulate elevated levels of anger during social interactions and solve social problems (Chorpita & Daleiden, 2009; Sukhodolsky et al., 2016). We designed two versions of YourSkills with identical content, but with different practice modes: one using virtual reality and one using roleplay. As clinicians often have to decide under uncertainty which treatment is most likely to be effective, we used Bayes Factors to indicate how likely it was that the virtual reality led to larger decreases in aggressive behavior compared to the comparison groups.

We conducted a randomized controlled trial with three conditions, comparing YourSkills virtual reality to YourSkills roleplay and care-as-usual. The first aim of our study was to examine treatment effects on children's aggressive behavior problems. As pre-registered in the clinical trial register, our first primary outcome measure was children's aggression. Specifically, we hypothesized that aggression decreases were larger for (1a) the two YourSkills groups versus the care-as-usual group, (1b) the YourSkills virtual reality versus the YourSkills roleplay group, and (1c) the YourSkills virtual reality versus the care-as-usual group. The second aim of our study was to investigate the potential experienced benefits of virtual reality above roleplays as treatment method for children with aggressive behavior problems. We hypothesized that children participating in YourSkills virtual reality would score higher than children participating in YourSkills roleplays on (2a) emotional engagement, (2b) practice immersion, (2c) treatment appreciation, and (2d) perceived efficacy. Our primary outcome was treatment appreciation (called treatment motivation in the pre-registration). Given that children's aggression and treatment appreciation were pre-registered as main

outcomes, their analysis should be considered as confirmatory. The other measures were later added to explore a broader range of potential advantages of virtual reality. Although these outcomes were based on previous literature and were planned in advance, they should be considered more exploratory as they were not pre-registered.

## METHOD

### Design

This study was a multicenter randomized controlled trial with three groups: YourSkills virtual reality, YourSkills roleplay, and care-as-usual. Children were recruited at fifteen clinical centers in the Netherlands providing mental health care for children with problems that are so severe that daily functioning is impaired and treatment is necessary. Recruitment began in September 2019, and all post-intervention assessments were completed by July 2021. Children were randomized at the individual level using computer-generated general random numbers. Specifically, we conducted randomization per clinical center. This study was approved by the Ethics Committee of the University Medical Centre Utrecht and was registered in the Dutch Trial Register (NTR; <https://trialsearch.who.int/Trial2.aspx?TrialID=NL7959>).

### Participants

Therapists working in the clinical centers were asked to approach parents of boys whose casefiles met our study's inclusion criteria: age 8–13 years, referred for displaying aggressive behavior problems, estimated intelligence level above 80, no severe autism spectrum disorder, and no epilepsy or severe visual or auditory limitations. Only boys were included, as aggression by girls in middle childhood may differ from aggression by boys in its development, processes, and outcomes (Berkout et al., 2011; Fontaine et al., 2009; Underwood, 2002). Moreover, the intervention with girls would require different virtual reality stimuli (e.g., girl avatars), which would be more feasible after a first 'proof of principal' with only boys in the present study. Children with severe autism spectrum disorder and low intelligence level were excluded, because Cognitive Behavioral Therapy exercises require perspective taking and imagination skills, as well as cognitive skills to reflect on thoughts and behavior (Sukhodolsky et al., 2016). We chose to exclude children with epilepsy and severe visual or auditory limitations as practicing in virtual reality would not be possible for them.

Consent was obtained for 127 children. Twelve children were excluded prior to intervention (for reasons, see Figure 1). Thus, 115 boys in the age range 8–13 years ( $M=10.58$ ,  $SD=1.44$ ) were included in the study. In addition to aggressive behavior problems at baseline (see

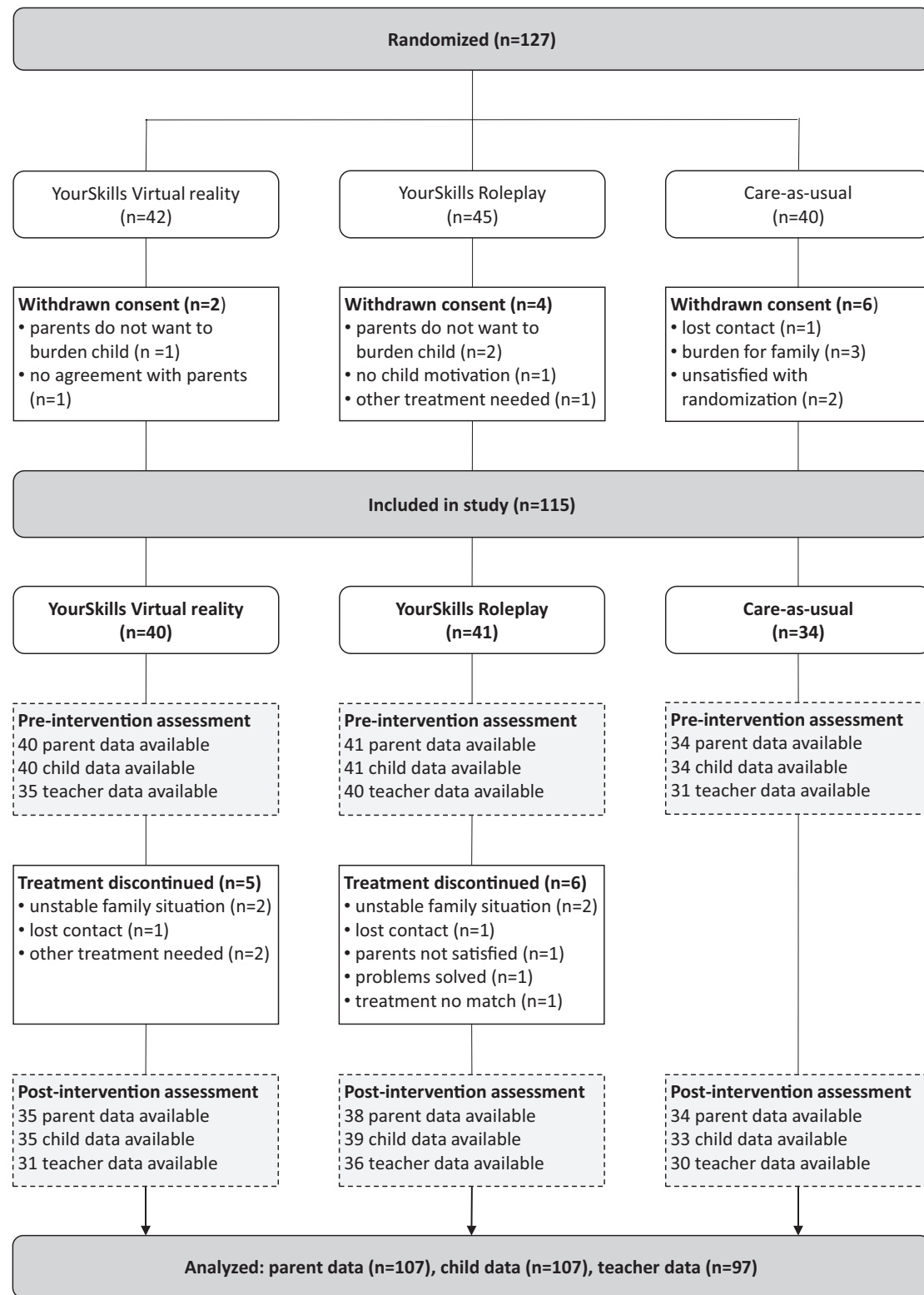
Table 2), we asked parents about children's diagnostic classifications, based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM), assessed by therapists when children entered the mental health care center. In our sample, 38 children (33.0%) were not classified with a disorder, 59 children (51.3%) were classified with one disorder, 16 children (13.9%) with two disorders, and two children (1.7%) with three disorders. Diagnoses included attention-deficit hyperactivity disorder ( $n=59$ ), oppositional deficit disorder ( $n=16$ ), autism spectrum disorder ( $n=14$ ), anxiety disorder ( $n=2$ ), attachment disorder ( $n=2$ ), and depressive disorder ( $n=1$ ). Children's intelligence level was on average 97.80 ( $SD=12.18$ ). Most children and parents were born in the Netherlands and most parents attained middle levels of education (see Table 1). After randomization, 40 children were assigned to the YourSkills virtual reality group, 41 to the YourSkills roleplay group, and 34 to the care-as-usual group.

Written informed consent was obtained from parents and 12- and 13-year old children. Participation was voluntary and children and parents were assured of confidential use of their data. Children received a small gift (e.g., a multicolor pen) after filling out the post-assessment. We also asked parents' consent to approach children's teachers to complete questionnaires (94.8% consent).

### Procedure

After randomization, therapy sessions were planned by the therapists, who then invited researchers to conduct the pre-assessment 30 min before the first therapy session. Researchers were invited again at the last therapy session to conduct the post-assessment directly after this session ended. When children in the care-as-usual group did not receive therapy at the clinical centers during the study, researchers planned home visits to conduct the pre- and post-assessments after randomization and 12 weeks later (i.e., the estimated average time of the YourSkills intervention). Children who discontinued the treatment ( $n=9$ ) were invited to remain in the study so that we could conduct intention-to-treat analyses and overcome problems with missing data (White et al., 2011).

All assessments with children were conducted face-to-face. Children were individually interviewed in 20–30 min by the first author or a trained research assistant. At the same time, parents were also asked to fill out questionnaires in an online system. When both parents were present during the assessment, we asked them to each fill out the questionnaires. For data analyses, we matched pre- and post-assessments of the same parent: mothers (45.2%) or fathers (27.0%). If both parents filled out both assessments (5.2%), we chose mother reports to align with the largest group filling out both assessments. In some cases, we had to combine mother- and father-report (20.9%) or had only pre-assessment data available



**FIGURE 1** Participant flow diagram.

(6.9%). We also asked teachers to fill out the pre- and post-assessment via an online questionnaire, in the same weeks as children and parents (82.6% provided both assessments; 9.6% only the pre-assessment; 1.7% only the post-assessment).

### YourSkills treatment

YourSkills is a manualized CBT, based on evidence-based treatments for children with aggressive behavior problems, including Coping Power (Lochman et al., 2008)

**TABLE 1** Background characteristics of participants included in the study.

	Total ( <i>n</i> =115)	YourSkills virtual reality ( <i>n</i> =40)	YourSkills roleplay ( <i>n</i> =41)	Care-as-usual ( <i>n</i> =34)
Age (years)	10.58 (1.44)	10.78 (1.49)	10.48 (1.56)	10.47 (1.21)
Intelligence level	97.80 (12.18)	95.03 (11.19)	98.85 (12.41)	99.62 (12.76)
Child born in the Netherlands	95.7%	95.0%	95.1%	97.1%
Parents born				
Both in the Netherlands	71.3%	72.5%	73.2%	67.6%
One in the Netherlands	14.8%	17.5%	12.2%	14.7%
Both elsewhere	13.9%	10.0%	14.6%	17.6%
Parental educational level				
Low education (ISCED 0–2)	18.3%	20.0%	19.5%	14.7%
Middle education (ISCED 3–4)	43.5%	47.5%	39.0%	44.1%
High education (ISCED 5–6)	38.2%	32.5%	41.5%	41.2%
Weeks pre- to post-assessment				
Parent- and child-reports	16.85 (7.98)	17.60 (7.33)	18.32 (10.41)	14.44 (4.36)
Teacher-reports	17.79 (9.26)	18.30 (8.15)	19.17 (11.36)	15.55 (7.07)
COVID-19 lockdown				
Finished before lockdown	9.6%	12.5%	2.4%	14.7%
In lockdown during study	18.3%	17.5%	19.5%	17.6%
Started after lockdown	72.2%	70.0%	78.0%	67.6%

Abbreviations: ISCED, International Standard Classification of Education (UNESCO, 2012).

and Self-Control (van Manen, 2001). We developed a new treatment manual, rather than adding virtual reality to an existing treatment. This way, we could integrate interactive virtual reality into all facets of the treatment and compare it to the identical treatment using roleplay practice. The aim of YourSkills is to reduce children's aggressive behavior problems by enhancing emotion regulation and social information processing skills. Children practice anger recognition, anger regulation, and social problem solving in social interactions. YourSkills consists of one 45-min introduction session with parents and ten 45-min sessions with the child (for an overview of the sessions, see Alsem et al., 2021). All treatment sessions have the same structure, making the session course predictable for children. Although YourSkills is primarily focused on the child, it also promotes parent involvement by providing them with an introduction session and including them at the end of each session (for more information, see Alsem et al., 2021).

To let children practice their regulation skills whilst being emotionally engaged, therapists create challenging social situations for children in virtual reality or roleplays. In each session, therapists first explain a new skill, then model the skill using roleplay, and then use virtual reality or roleplays to let children practice the skill in anger-provoking social situations. The YourSkills materials include 26 cards with anger-provoking situations, based on a taxonomy of problematic situations for children with aggressive behavior problems. They include: being disadvantaged, authority conflicts, peer rejection,

and peer provocation (Matthys et al., 2001). Therapists select those situations that match children's individual needs.

In this study, YourSkills was delivered by 31 licensed therapists (90.3% female) working at the participating clinical centers. All therapists had experience providing treatment to children and adolescents, ranging from 2.0 to 25.1 years ( $M=7.79$ ,  $SD=5.75$ ). Therapists' experience with treatments specifically for children ages 8–13 years with aggressive behavior problems ranged from 0.5 to 12.5 years ( $M=5.72$ ,  $SD=4.07$ ), with only one therapist having less than 1 year of experience. Most therapists had completed a post-master course in Cognitive Behavioral Therapy (87.1%). Therapists were trained in both versions of YourSkills in a two-day course, supervised by the first and second author and a certified CBT therapist. They learned to work with the treatment manual, how to conduct roleplay exercises, and use the virtual reality equipment. These therapists used the same treatment manual for both versions of YourSkills, and only distinguished in practice mode during the exercises by using either virtual reality or roleplays depending on the condition their client was assigned to. Thus, therapist characteristics were equal across conditions. As the only difference between the two treatment versions was the practice mode, contamination of one version to the other was not likely. During the treatment period, therapists could receive consultation over the phone from the first or second author. The focus of the consultation was on help with practical issues, rather than supervision. Few



**TABLE 2** Pre-intervention assessment and post-intervention assessment means (*M*) and standard deviations (*SD*) of the outcome variables for the YourSkills virtual reality group, YourSkills roleplay group, and the care-as-usual group.

	YourSkills virtual reality				YourSkills roleplay				Care-as-usual			
	Pre-test		Post-test		Pre-test		Post-test		Pre-test		Post-test	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Aggression frequency parents (IRPA)	1.91	0.59	1.70	0.43	2.12	0.64	1.84	0.63	1.93	0.66	1.91	0.58
Aggression frequency teacher (IRPA)	2.03	0.70	1.77	0.63	2.04	0.88	1.93	0.83	1.99	0.71	1.92	0.72
Aggression frequency child (IRPA)	1.82	0.56	1.67	0.48	2.00	0.69	2.04	0.71	1.97	0.74	1.82	0.63
Weekly aggression parents	2.58	0.74	1.78	0.54	2.57	0.90	2.29	0.78	2.48	0.86	2.38	0.80
Weekly aggression child	1.99	0.94	1.62	0.57	2.14	1.15	1.77	0.67	2.10	1.01	2.28	1.10
Aggressive behavior parents (CBCL)	15.91	6.30	11.83	4.71	17.63	6.80	13.90	6.80	16.94	7.27	13.82	5.71
Aggressive behavior teacher (TRF)	17.37	10.35	13.77	10.24	17.17	10.44	14.61	10.67	16.38	10.89	14.45	10.45

Abbreviations: CBCL, Child Behavior Checklist; IRPA, Instrument for Reactive and Proactive Aggression, TRF, Teacher Report Form.

therapists used the opportunity for consultation, and most questions concerned exclusion criteria for study participation or technical questions about the virtual reality equipment (e.g., the laptop is not starting).

Therapists managed to carry out almost all session elements of YourSkills (virtual reality:  $M=98.6\%$ ; roleplay:  $97.1\%$ ; for a session description, see Alsem et al., 2021). Therapists indicated that children practiced more than the recommended 10 min per session (virtual reality:  $M=11.8$  min,  $SD=2.2$ ; roleplay:  $M=11.4$ ,  $SD=2.1$ ). Within this practice time, children practiced their new skill more often than the recommended two times (virtual reality:  $M=3.0$ ,  $SD=0.7$ ; roleplay:  $M=3.3$ ,  $SD=1.0$ ). Therapists were satisfied with how they delivered the treatment (virtual reality:  $M=4.2$  on a 5-point scale,  $SD=0.6$ ; roleplay:  $M=4.3$ ,  $SD=0.6$ ).

### YourSkills virtual reality

The YourSkills virtual reality software includes practice scenarios that correspond with the YourSkills cards describing anger-provoking situations. The virtual reality environment consists of a classroom, a schoolyard, and a living room (for an impression, see Figure 2). Children wore an Oculus Rift S headset, a noise canceling headphone, and they held controllers in both hands, allowing them to grab and throw virtual objects. In the first session, therapists explained to children that the virtual environment allowed them to walk around freely (within a  $3 \times 3$  meter area), talk with virtual children and adults, and play games such as building a tower or playing a game on the television. Therapists could evoke children's anger by manipulating the virtual situation itself (e.g., letting the child lose a game, or switching off the television) or by manipulating the speech and actions of the virtual characters. Therapists used a microphone with voice transformer to emulate a different voice for each virtual character. They used a tablet to control the characters' bodily movements (e.g., walking away), gestures

(e.g., raising a middle finger), and facial expressions (i.e., an expression scale from happy to angry).

### YourSkills roleplay

The YourSkills roleplay version was identical to the virtual reality version, except that children did not practice in virtual reality but in roleplays. Therapists used the cards describing anger-provoking situations to roleplay challenging social situations, and played the role of a child's parent, teacher or peer. Therapists were encouraged to use physical objects and make use of the room to stimulate active engagement of children during the roleplays.

### Care-as-usual

Children in the care-as-usual group received the usual care provided by the clinical institutions. Trained therapists for this study were not allowed to provide care-as-usual to this group, to assure that they did not make use of YourSkills' treatment elements. We expected a variety of care (Kazdin, 2015), including individual therapy, group therapy, and parent training. At post-assessment, we asked parents to fill out whether and what therapy they or their children received for children's aggressive behavior problems.

### Treatment participation in routine care

As children were recruited in routine care, children in all intervention groups were allowed to receive other services when needed. Few children in the YourSkills groups participated in additional treatments: Some children also received medication (virtual reality:  $n=6$ ; roleplay:  $n=4$ ) and some parents also participated in parent training for the aggressive behavior problems of their child (virtual reality:  $n=4$ ; roleplay:  $n=2$ ).



FIGURE 2 Virtual reality classroom, living room and schoolyard environments.

During the study period, 50% of the 34 families in the care-as-usual group participated in treatments specifically aimed at decreasing children's aggressive behavior problems. The other families indicated that they did not participate in a treatment specifically aimed at these problems. Of the 17 children participating in routine care, 14 participated in some form of individual therapy covering on average 9.3 sessions ( $SD=5.4$ ). Of these 14 children, three also received medication and one family participated in parent training. Of the other 3 children, one child participated in five group sessions, and one child participated in two group sessions, received medication, and his parents participated in a training. In one family only parents participated in parent training.

## Measures

We here present the measures assessed to answer this studies' research questions. We assessed additional measures for other purposes, which are not reported here.

### Children's aggressive behavior

To obtain a comprehensive picture of changes in children's aggression, we used a multi-informant (parent-, child-, and teacher-reports), multi-instrument approach. Including multiple informants is highly informative as aggressive behavior is context-dependent and the

correspondence between informants is relatively low (De Los Reyes et al., 2015). We used three instruments, providing different information on children's aggressive behavior. First, we assessed aggressive behavior in the past month using a widely used instrument (i.e., the ASEBA forms: CBCL and TRF; Achenbach & Rescorla, 2001). This instrument has normative data for parent- and teacher-report, allowing us to investigate changes from clinical to normative levels of aggressive behavior. Second, we assessed the frequency of aggression in the past month with the validated IRPA questionnaire (Polman et al., 2009). This instrument is not only suitable for parent- and teacher-report but also for child-report and may be more sensitive to small changes in behavior as it uses a 5-point scale (instead of the 3-point scale in the ASEBA forms). Third, we included a new weekly report measure assessing children's aggression in the past week (Alsem et al., 2022), allowing us to capture short-term changes in aggression, directly after the intervention ended.

### CBCL and TRF aggressive behavior

Parents and teachers filled out the aggressive behavior scale of the Child Behavior Checklist (CBCL) and the Teacher Report Form (TRF), respectively (Achenbach & Rescorla, 2001). They rated children's aggressive behavior in the past month on a 3-point scale (0=not true, 1=somewhat true, 2=very true or often true). The CBCL scale consists of 18 items (e.g., "Argues a lot") and the TRF scale of 20 items (e.g., "Physically attacks people"). We used norms for Dutch children to calculate T-scores

to examine (sub)clinical levels of aggression, and calculated sum scores for all other analyses. In our sample, the internal consistency was adequate for both parents and teachers at pre- and post-assessment ( $\alpha$  .86–.95).

#### *IRPA aggression frequency*

Parents, teachers, and children filled out the Instrument for Reactive and Proactive Aggression (IRPA; Polman et al., 2009). They rated the frequency of aggression in the past month on 7 items (e.g., “How often did your child/the child/you hit someone in the past month?”) on a scale from 1 (never) to 5 (daily). Ratings were averaged across items, with adequate internal consistency for all informants at pre- and post-assessment ( $\alpha$  .74–.86).

#### *Weekly report measure*

Parents and children filled out a weekly report measure (Alsem et al., 2022). They rated three items (e.g., “This week my child/I fought with someone”) on a scale from 1 (never) to 5 (very often). Ratings were averaged across items. The child-report version showed adequate internal consistency, convergent, and concurrent validity in a previous study (Alsem et al., 2022). The internal consistency in the current study was adequate for both parents and children at pre- and post-assessment ( $\alpha$  .75–.79).

### Measures assessing the potential benefits of virtual reality

To investigate the potential benefits of virtual reality as treatment method for children with aggressive behavior problems, children and parents rated items about their experience with YourSkills at post-assessment. Therapists filled out items about the two versions of YourSkills after the study ended (we counterbalanced the order of items on virtual reality versus roleplay).

#### *Emotional engagement*

Children and therapists rated children's emotional engagement while practicing in the virtual reality or roleplays on three items (i.e., “Some things in the virtual reality/roleplays really pissed me/the children off a bit,” “I/the children never felt anger in the virtual reality,” and “Sometimes I/the children felt like getting angry in the virtual reality”) on a scale from 1 (totally disagree) to 5 (totally agree). Ratings were averaged across items, with adequate internal consistency for both children and therapists ( $\alpha$  .77–.91).

#### *Practice immersion*

Children and therapists rated four items on practice immersion during virtual reality or roleplays (i.e., “I/the children was/were completely immersed in virtual reality/the roleplays,” “The virtual reality felt real (for the kids),” “I felt/the children were feeling like the virtual reality really happened to me/them,” and “During the

virtual reality it felt like I/the children was/were actually experiencing it”) on a scale from 1 (totally disagree) to 5 (totally agree). Ratings were averaged across items, with adequate internal consistency for children and therapists ( $\alpha$  .84–.88).

#### *Treatment appreciation*

Children, parents, and therapists rated four items about children's treatment appreciation (e.g., “I/my child/the children liked to participate in YourSkills”) on a scale from 1 (totally disagree) to 5 (totally agree). Ratings were averaged across items. The internal consistencies were adequate for parents and children ( $\alpha$  .80–.89) and the therapist roleplay scale ( $\alpha$  = .90) but not for the therapist virtual reality scale ( $\alpha$  = .59). To gain an overall impression of children's appreciation of YourSkills, we also asked children give a grade from 1 to 10 to the treatment as a whole and to practicing in the virtual reality/roleplays.

#### *Perceived efficacy*

Children, parents, and therapists rated four items on their perceived efficacy of the treatment (e.g., “I/my child/the children learned a lot in YourSkills”) on a scale from 1 (totally disagree) to 5 (totally agree). Ratings were averaged across items. The internal consistencies were adequate for children ( $\alpha$  = .83) and parents ( $\alpha$  = .76) and the therapist roleplay scale ( $\alpha$  = .75), but not the therapist virtual reality scale ( $\alpha$  = .54).

### Intelligence

When information on intelligence was available from children's casefile (administered within the past 2 years; 59.1% of the cases), we did not assess intelligence again. Otherwise, we administered the subtests ‘Block Design’ and ‘Vocabulary’ of the Wechsler Intelligence Scale for Children (WISC-III; Kort et al., 2005) to estimate an IQ score (Silverstein, 1970). Such estimated IQ scores are strongly associated with IQ scores based on the total WISC (Hrabok et al., 2014).

### Analyses

We conducted our analyses using Bayesian statistics, a method that is becoming more common in social and behavioral sciences (van de Schoot et al., 2014). An advantage of Bayesian statistics is that it quantifies the amount of support for the study hypotheses instead of yielding a dichotomous decision on whether the null hypothesis is rejected or not (van de Schoot et al., 2014). This provides clinicians with an indication of which treatment is most likely to be effective. Another reason to use Bayesian analyses was to overcome problems with our large number of outcome measures. Specifically, a



major advantage of Bayesian analyses is that there are no risks for type I or type II errors when conducting multiple analyses (Hojtink et al., 2019). Moreover, our sample size was smaller than the intended sample size, due to COVID-19 related inclusion problems. As we did not specify the analytic approach beforehand, we chose to use Bayesian statistics to minimize problems with our smaller sample size.

Bayesian analyses yield Bayes factors (BF), which quantify to what extent the data support one hypothesis compared to another. A Bayes factor of 1 indicates equal support for both hypotheses; a Bayes factor of  $>1$  indicates support in favor of the planned hypothesis over the null hypothesis, with higher Bayes factors providing more support. For instance, if we would find  $BF=10$  for the hypothesis that YourSkills virtual reality leads to larger decreases in aggression than care-as-usual, this would indicate that it is 10 times more likely that YourSkills indeed outperformed care-as-usual than not. We conducted our statistical analyses in JASP version 0.15.0.0 with the Bain package (Hojtink et al., 2019; Marsman & Wagenmakers, 2017).

Before we statistically tested our hypotheses, we explored clinically relevant changes in aggression. We used the available norms of the CBCL and TRF to calculate T-scores and classify children in the normal range (T-score  $\leq 64$ ,  $\leq 93$ rd percentile), subclinical range (T-score 65–69, 94–97th percentile), or clinical range (T-score  $> 69$ ,  $> 97$ th percentile). We defined clinically relevant improvement as a shift from one range to another from pre- to post-assessment. Next, we preliminarily explored if there was an overall decrease in aggression across groups. We conducted Bayesian paired sample *t*-tests to test our prediction that post-intervention levels of aggression were lower than pre-intervention levels against the contrasting prediction that pre- and post-intervention aggression levels were equal. As our analysis included two parameters (i.e., mean difference and mean difference variance), we set the fraction on two (Hojtink et al., 2019).

To investigate our first research question, we tested whether decreases in aggression were larger for (1a) the two YourSkills groups versus the care-as-usual group, (1b) the YourSkills virtual reality versus the YourSkills roleplay group, and (1c) the YourSkills virtual reality versus the care-as-usual group. We first calculated mean difference scores by subtracting the pre-intervention from the post-intervention scores. We then conducted Bayesian ANOVA's, to test the hypothesis that the mean difference scores differed between the groups in expected directions against the complementary hypothesis (e.g., for hypothesis 1c: that the mean differences on aggression were *not* larger in the YourSkills virtual reality versus care-as-usual group). In addition, we calculated Cohen's *d* effect sizes based on the means and standard deviations (Cohen, 1988).

To investigate our second research question, examining the potential benefits of virtual reality versus roleplays,

we used Bayesian one-way ANOVAs to analyze whether children practicing in virtual reality showed higher levels of (2a) treatment appreciation, (2b) emotional engagement, (2c) practice immersion, and (2d) perceived efficacy. Again, each hypothesis was tested against its complement. We also calculated Cohen's *d* effect sizes (Cohen, 1988).

To check for missing data patterns, we conducted Little's test, which produced a normed  $\chi^2$  (i.e.,  $\chi^2/df$ ) of 1.16. Thus, our data did not refute the null hypothesis that the data were missing completely at random (Bollen, 1989). Therefore, we used default settings in JASP (i.e., listwise deletion). This means that participants who did not fill out post-assessment were excluded from the analyses. We tried to avoid exclusion by asking participants to remain in the study after drop-out, to be able to conduct intention-to analyses and overcome problems with missing data due to dropout (White et al., 2011). In the intention-to-treat principle, all randomized participants are included in the analyses in the groups to which they were randomized, even if they stopped treatment early. This method is preferred in randomized trials as these analyses give unbiased, conservative estimates of treatment effects, and allow for the greatest generalizability (Gupta, 2011). In total, we analyzed data of 107 children and parents (7.0% missing) and 97 teachers (15.7% missing; see Figure 1). To check whether listwise deletion may have biased our results, we conducted a robustness check using single imputation for our aggression measures (note that multiple imputation is not possible within JASP). Conclusions from these analyses were the same as reported here (see Supplementary Material).

## RESULTS

### Preliminary analyses

#### Pre-intervention group differences

To check whether randomization was successful, we examined between-group differences at pre-assessment. Results showed that it was more likely that there were *no* group differences in background characteristics (Table 1) than that there *were* group differences, with BFs favoring no differences ranging from 3.04 to 128.49. Next, we compared pre-intervention levels of aggression (Table 2), and found that it was more likely that groups did *not* differ than that they *did* differ at baseline, according to all aggression measures by all informants, with BFs ranging from 5.42 to 11.44. We also checked whether the three groups were differentially affected by the COVID-19-related lockdown (i.e., Dutch clinical institutions were closed from March 22th to May 11th 2019; Table 1). First, we inspected the number of children affected by the lockdown, and found that no differences were more likely,  $BF=36.62$ . Second, we inspected the time between pre- and post-assessment. Results showed no differences

were more likely, but only for the time between teacher-reports ( $BF=3.22$ ) and not for parent- and child-reports ( $BF=1.21$ ). Time between pre- and post-assessment for these reports was on average 14 weeks in the care-as-usual group, and 18 weeks in the two YourSkills groups.

### Descriptive clinical decreases in aggression (CBCL and TRF)

Figure 3 presents the average aggression T-scores at pre- and post-assessment for parent- and teacher-report. All three groups decreased in parent-reported aggression: from the subclinical to the normal range for the YourSkills virtual reality group, and from the clinical to the subclinical range for the other two groups. Teacher-reported aggression also decreased in all three groups, but all groups started and remained in the subclinical range.

We then explored percentages of children who did not change, improved or deteriorated (i.e., shifted from one range to another; Figure 4). For parent-reported aggression, most children improved in the YourSkills virtual reality group (48.6%), followed by the YourSkills roleplay group (39.5%), and care-as-usual group (26.5%). Many children remained in the same range, but the least in the YourSkills virtual reality group (42.9%), followed by the YourSkills roleplay group (50.0%) and the care-as-usual group (64.7%). Some children deteriorated (i.e., 8.6–10.5% across groups). For teacher-reported aggression, a slightly different pattern was found. Most children remained in the same range in all three groups (61.1–66.7%), whilst in the YourSkills roleplay group more children improved

(30.6%) than in the virtual reality and care-as-usual groups (23.3% and 17.2%, respectively). Deterioration was highest in the care-as-usual group (17.2%).

### Overall decreases in aggression

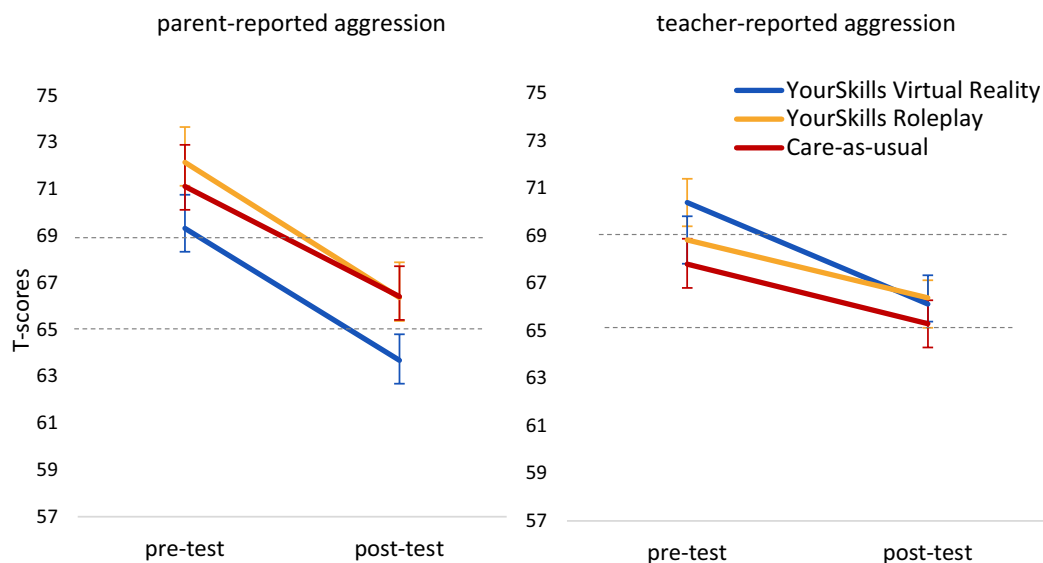
We used Bayesian paired sample *t*-tests to explore decreases in aggression from pre- to post-assessment across the three intervention groups. We found that decreases in aggression were more likely than no change for six out of seven aggression measures (Table 3). This was over 36 times more likely for all three parent-reported aggression measures and teacher-reported aggressive behavior (TRF) but only about two times more likely for teacher-reported aggression frequency (IRPA) and child-reported weekly aggression. We found no support for decreases in child-reported aggression frequency ( $BF<1$ ).

### Research question 1: Group differences in aggression decreases

To investigate group differences in aggression decreases from pre- to post-assessment (Figure 5), we conducted Bayesian ANOVAs.

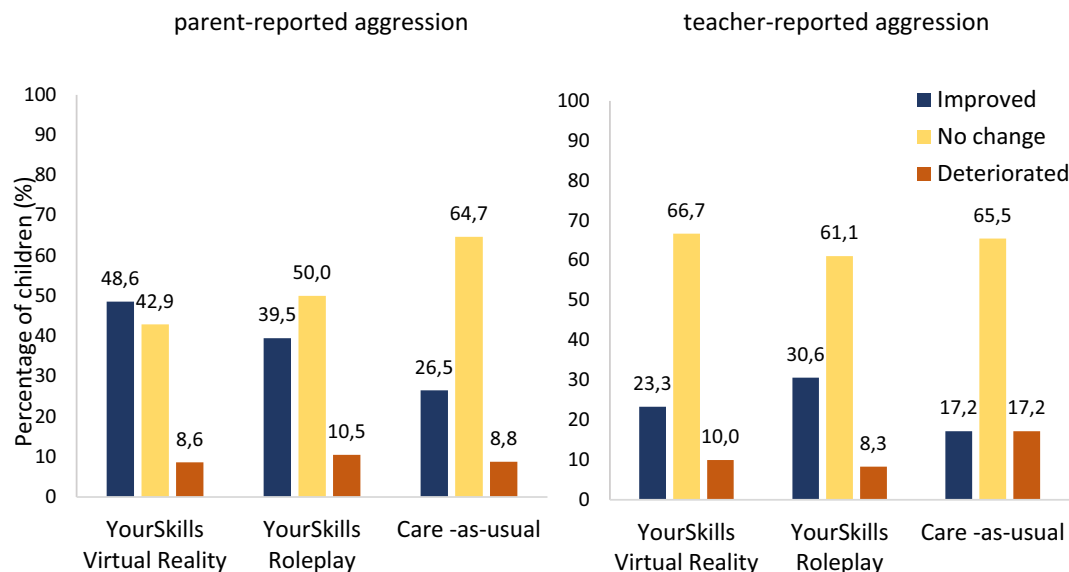
### YourSkills (both versions) versus care-as-usual

Six out of seven aggression measures suggested superior effectiveness of YourSkills compared to care-as-usual



Note. Error bars show standard errors. Dashed lines represent the subclinical (lower line) and clinical (upper line) T-score cut-offs.

FIGURE 3 Average T-scores for each group at pre- and post-assessment for both parent-reported aggression (CBCL; Left) and teacher-reported aggression (TRF; right).



**FIGURE 4** Percentages of children who improved, did not change, or deteriorated in each group for both parent-reported aggression (CBCL; Left) and teacher-reported aggression (TRF; right).

(Table 3). It was at least 187 times more likely that YourSkills outperformed care-as-usual than not according to parent- and child-reported weekly aggression ( $d_s=0.55$ ), 27 times more likely according to parent-reported aggression frequency (IRPA;  $d=0.39$ ), but only 3 times more likely according to teacher-reported aggression frequency (IRPA) and parent- and teacher-reported aggressive behavior (CBCL/TRF;  $d_s 0.14-0.16$ ). For child-reported aggression frequency (IRPA), it was more likely that the YourSkills groups did *not* improve more than the care-as-usual group ( $BF < 1$ ).

### Virtual reality versus roleplay

Results for four out of seven aggression measures favored virtual reality over roleplays (Table 3). It was 528 times more likely that virtual reality outperformed roleplay than not according to parent-reported weekly aggression ( $d=0.68$ ), but only 2–8 times more likely according to teacher-reports of aggression (IRPA and TRF) and child-reported aggression frequency (IRPA;  $d_s 0.14-0.30$ ). We found no support for larger aggression decreases in virtual reality relative to roleplay on the other parent-reported aggression measures (IRPA and CBCL) and weekly child-reported aggression ( $BFs < 1.53$ ,  $d_s < 0.06$ ).

### Virtual reality versus care-as-usual

Results for six out of seven aggression measures favored virtual reality over care-as-usual (Table 3). It was 84 times more likely that virtual reality outperformed care-as-usual than not, according to child- and parent-reported

weekly aggression ( $d_s 0.59-0.95$ ), and 3–10 times more likely according to other parent- and teacher-reports of aggression (IRPA and CBCL/TRF;  $d_s 0.19-0.33$ ). For child-reported aggression frequency (IRPA), we found no support for larger aggression decreases in virtual reality versus care-as-usual ( $BF < 1$ ).

### Research question 2: Additive value of virtual reality

We conducted Bayesian ANOVAs to analyze whether children in the virtual reality group experienced higher levels of emotional engagement, practice immersion, treatment appreciation, and perceived efficacy than children in the roleplay group.

### Emotional engagement and practice immersion

Results showed that it was very likely that children were more emotionally engaged during practice in virtual reality versus roleplays, as suggested by both child and therapist-report ( $BFs > 60.20$ ,  $d_s 0.58-0.60$ ; Table 4). For practice immersion, results indicated that it was very likely that children practicing in virtual reality felt more immersed than children practicing in roleplays, according to both child- and therapist-reports ( $BFs > 48.57$ ,  $d_s 0.48-1.05$ ).

### Treatment appreciation and perceived efficacy

Results showed that it was likely that children in the virtual reality group appreciated the treatment more

**TABLE 3** Bayes factors (BF) and Cohen's *d* effect sizes (*d*) for decreases in aggression and group comparisons on mean differences between pre-intervention and post-intervention assessments.

	Decrease aggression across groups		YourSkills (both versions) vs. care-as-usual		Virtual reality vs. roleplay		Virtual reality vs. care-as-usual	
	BF	<i>d</i> [95% CI]	BF	<i>d</i> [95% CI]	BF	<i>d</i> [95% CI]	BF	<i>d</i> [95% CI]
Aggression frequency parents (IRPA)	36.56	0.39 [-0.04, 0.81]	27.64	0.39 [-0.04, 0.81]	0.47	-0.12 [-0.58, 0.35]	9.95	0.31 [-0.17, 0.80]
Aggression frequency teacher (IRPA)	2.78	0.16 [-0.27, 0.60]	3.34	0.16 [-0.27, 0.60]	4.38	0.20 [-0.29, 0.68]	6.28	0.33 [-0.19, 0.84]
Aggression frequency child (IRPA)	0.66	-0.16 [-0.57, 0.25]	0.29	-0.16 [-0.57, 0.25]	8.38	0.30 [-0.16, 0.76]	0.95	<-0.01 [-0.48, 0.47]
Weekly aggression parents	68,280.95	0.55 [0.12, 0.97]	187.31	0.55 [0.12, 0.97]	528.82	0.68 [0.19, 1.15]	108.77e <sup>2</sup>	0.95 [0.43, 1.45]
Weekly aggression child	1.99	0.55 [0.13, 0.96]	223.20	0.55 [0.13, 0.96]	1.03	<0.01 [-0.45, 0.46]	84.38	0.58 [0.09, 1.06]
Aggressive behavior parents (CBCL)	173.60e <sup>7</sup>	0.14 [-0.27, 0.55]	3.01	0.14 [-0.27, 0.55]	1.53	0.06 [-0.40, 0.52]	3.21	0.19 [-0.29, 0.66]
Aggressive behavior teacher (TRF)	322.92	0.16 [-0.28, 0.59]	3.14	0.16 [-0.28, 0.59]	2.64	0.14 [-0.35, 0.62]	4.50	0.24 [-0.27, 0.75]

Note: A Bayes factor of >1 indicates support of our hypothesis (i.e., decreased aggression).

Note: e is used for readability and indicates exponential notation, with the corresponding number indicating how many times the BF must be multiplied by ten to generate the original number.

Abbreviations: CBCL, Child Behavior Checklist; CI, confidence interval; IRPA, Instrument for Reactive and Proactive Aggression; TRF, Teacher Report Form.

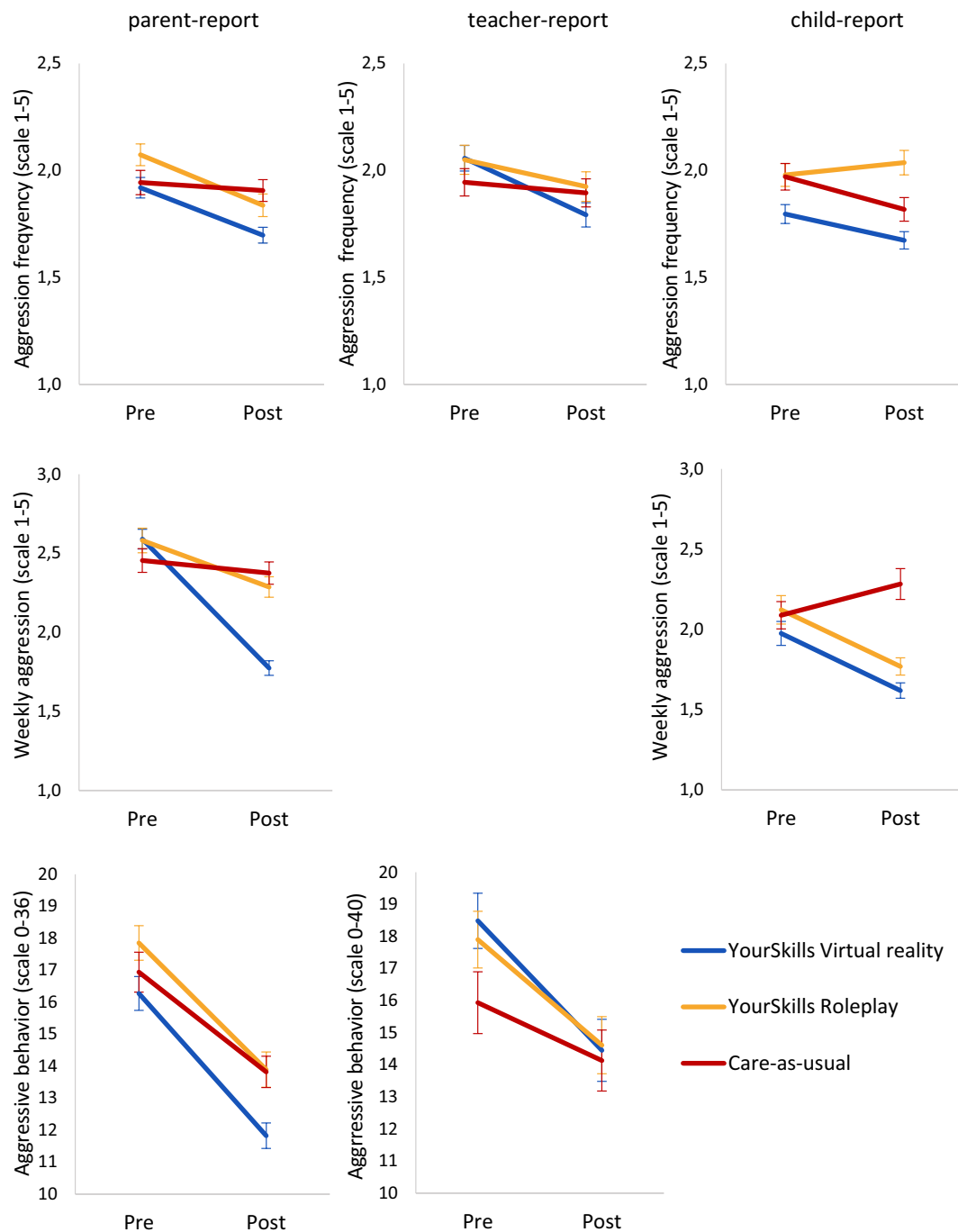
than children in the roleplay group, according to themselves (BF = 13.93,  $d=0.35$ ), their therapists (BF > 100,000,  $d=1.23$ ) and their parents (BFs = 79.23,  $d=0.54$ ; Table 4). Further, results showed that it was only somewhat more likely that children participating in the virtual reality version rated the treatment overall with a higher grade than children in the roleplay version (BF = 2.71,  $d=0.14$ ), but much more likely that children rated virtual reality as practicing method with a higher grade than roleplays (BF = 1631.37,  $d=0.75$ ). For perceived efficacy, reports showed that it was much more likely that therapists perceived virtual reality as more effective than roleplays (BFs = 568.47,  $d=0.62$ ; Table 4), and only somewhat more likely that children and parents had this perception (BFs 2.40–2.65;  $ds$  0.13–0.14).

## DISCUSSION

The present multicenter randomized controlled trial examined whether interactive virtual reality enhanced the effectiveness of CBT for boys with aggressive behavior problems compared to CBT with roleplays and care-as-usual. The results indicated that CBT with virtual reality was more likely to reduce aggressive behavior than care-as-usual for six out of seven outcomes. Effects were medium-to-large for measures assessing weekly aggression ( $ds$  .59–.95) and small-to-medium for measures assessing aggression in the past month ( $ds$  .19–.33). The same pattern of results was found when we compared both CBT groups (i.e., virtual reality and roleplays) to care-as-usual, suggesting that our newly developed CBT protocol outperformed care-as-usual. When we directly compared virtual reality versus roleplays, results favored virtual reality on four of seven aggression measures, with small-to-medium effect sizes ( $ds$  .14–.68). Virtual reality clearly outperformed roleplays on other aspects: it was very likely that children were more emotionally engaged and immersed during virtual reality practice than in roleplays. Also, children most likely appreciated virtual reality more and perceived this method as more effective than roleplays.

Our findings provide the first indication that interactive virtual reality can enhance effects of CBT for children with aggressive behavior problems. Effect sizes for virtual reality versus care-as-usual were substantial ( $ds$  .19–.95) and similar or larger than in meta-analytic research comparing CBT to control groups ( $d=.23$ ; McCart et al., 2006). In line with these effects, 48.6% of parents in the CBT virtual reality group reported clinically relevant improvements in children's aggression, and parent-rated average aggression scores decreased from subclinical levels to the normal range. Moreover, virtual reality likely enhanced children's treatment appreciation and involvement. This is highly relevant, as children with aggressive behavior problems are often not motivated for treatment (Frick, 2012; Lochman et al., 2019), whereas





Note. Error bars show standard errors

**FIGURE 5** Pre- to post-intervention aggression reports of parents, teachers, and children for the YourSkills virtual reality group, YourSkills roleplay group and the care-as-usual group.

enhancing treatment appreciation has been related to increases in treatment effectiveness (Lochman, Kassing, & Sallee, 2017).

Interactive virtual reality had some benefits over CBT with roleplays. Children practicing in virtual reality were more emotionally engaged and immersed, and we found some indications that virtual reality outperformed roleplays in effectiveness. These findings align with the dual-mode social information processing model for children with aggressive behavior problems (Verhoef et al., 2022).

This model proposes that children process social information in either the automatic mode (i.e., fast, emotion-driven aggression) or the reflective mode (i.e., slow, deliberately selected aggression). Based on this model, interventions may be most effective when children's social information processing patterns are targeted in the mode that is also active when they engage in aggressive behavior in daily life. Virtual reality may trigger the automatic mode more so than roleplays, as children practice in realistic environments and do not have to rely on

**TABLE 4** Means (*M*), standard deviations (*SD*), Bayes factors (*BF*) and Cohen's *d* effect sizes (*d*) of emotional engagement, practice immersion, treatment appreciation, and perceived efficacy.

	YourSkills virtual reality		YourSkills roleplay		Virtual reality vs. roleplay	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>BF</i>	<i>d</i> [95% <i>CI</i> ]
Engagement						
Child	2.77	1.15	2.13	1.08	153.43	0.58 [0.11, 1.04]
Therapist	3.71	1.16	3.06	0.96	60.20	0.60 [0.01, 1.18]
Immersion						
Child	3.35	1.05	2.81	1.18	48.57	0.48 [0.01, 0.94]
Therapist	3.54	0.93	2.58	0.89	71,293.20	1.05 [0.43, 1.65]
Appreciation						
Child	4.24	0.96	3.87	1.15	13.93	0.35 [-0.11, 0.81]
Therapist	4.56	0.45	3.69	0.91	176,419.75	1.23 [0.59, 1.83]
Parent	4.35	0.86	3.86	0.97	79.23	0.54 [0.05, 1.01]
Efficacy						
Child	4.36	0.65	3.79	0.73	2.65	0.14 [-0.32, 0.60]
Therapist	4.24	0.59	3.83	0.74	568.47	0.62 [0.02, 1.19]
Parent	4.22	0.90	4.11	0.79	2.40	0.13 [-0.34, 0.60]
Rating						
YourSkills	8.40	1.75	8.10	2.34	2.71	0.14 [-0.32, 0.60]
Method	8.66	1.51	6.90	2.88	1631.37	0.75 [0.27, 1.22]

Abbreviations: *CI*, confidence interval.

their memory or imagination (Park et al., 2011), triggering the reflective mode.

Although our results provide first indications that both virtual reality and our newly developed CBT protocol outperformed care-as-usual, we did find marked differences between outcome measures. Effects on measures assessing aggression in the past *month* (i.e., CBCL/TRF and IRPA) were generally smaller than effects on measures assessing aggression in the past *week* (i.e., weekly aggression measure). We propose three explanations for this discrepancy. First, we used measures that were validated to assess children's aggression in the past month. However, at post-assessment, this month included the last few weeks of the treatment period. In these weeks children still needed to learn new skills and generalize these to daily life, and so the monthly measures may have underestimated treatment effects. Second, the weekly measures might have been more sensitive to short-term changes in behavior as these items were specifically developed to capture this (Asem et al., 2022). Third, questions concerning a longer time period may be more strongly affected by the 'halo effect': a generalized impression of a child as 'aggressive' (Abikoff et al., 1993). The weekly measures may have been less susceptible to negative views that parents and teachers may have developed about children with behavior problems (DeVries et al., 2017).

Effects of CBT with virtual reality also differed between informants. Child-reported effects on aggression were generally smaller than effects reported by parents

and teachers. One explanation is that children may have underreported their aggressive behavior problems at pre-assessment (e.g., due to external attributions of their own behavior; De Los Reyes & Kazdin, 2005). Children may then have become more aware of their problems during the treatment (i.e., response shift bias; Rioux & Little, 2020), which is in line with our finding that we found little support for decreases in child-reported aggression across all treatment groups. Alternatively, parents and teachers may have overreported effects of the treatment. They were not blind to allocation status, and may have expected the novel virtual reality treatment to be more effective. However, this alternative explanation seems less likely, as intervention effects on parent reported measures have generally been found to be similar in magnitude to actual observed effects (Menting et al., 2013).

Strengths of this study include the randomized design with two comparison groups that allowed us to compare virtual reality with care-as-usual and examine the additive value of virtual reality compared to roleplays. We included multiple clinical centers, and recruited children in routine care. We used a multi-informant approach, which is highly informative as aggressive behavior is context-dependent and the correspondence between informants is relatively low (De Los Reyes et al., 2015). Last, we used Bayesian statistics, presenting the enhanced effectiveness of virtual reality in terms of likelihood, which is relevant for clinicians who have to decide under uncertainty which treatment is most likely to be effective.

Our study also had its limitations. First, we were not able to achieve our preregistered sample size, which was inevitable given the COVID-19 situation. We used Bayesian statistics, which are still influenced by smaller sample sizes (i.e., lower Bayes factors reflect less certainty), but allowed us to quantify the amount of support for our hypotheses instead of yielding a dichotomous decision based on an arbitrary cut-off (e.g.,  $p < .05$ ; Cohen, 1988) that may have been unduly influenced by a lack of power (van de Schoot et al., 2014). Second, only half of the families in the care-as-usual group participated in treatment for aggressive behavior problems during the study (although they could receive treatment for other problems). This group should thus be seen as a partly passive control group, and effects might have been smaller when care-as-usual treatment participation had been higher. On the other hand, our control group does reflect treatment received by children in routine care in clinical centers in the Netherlands. Third, we were not able to obtain information about the therapists that provided care-as-usual within the clinical institutions. As such, we were unable to examine whether care-as-usual therapists had the same level of experience and training as the therapists providing the YourSkills treatment. Although therapists in both conditions came from the same clinical institutions, it is possible that therapists signing up to provide YourSkills differed from the ones who did not (e.g., more enthusiasm or experience). Fourth, due to ethical and practical regulations in the clinical centers, pre-assessments were conducted after randomization. Parents, children and teachers were not blind to allocation during the pre-assessment, which might in theory have influenced their reports of children's aggressive behavior problems, as well as drop-out rates. Specifically, our randomization procedure could not prevent us from ending up with unequal sample sizes over conditions. The care-as-usual group was smallest, due to the highest number of parents who withdrew consent after randomization ( $n=6$ ). This was inevitable, given the ethical requirement that consent can be withdrawn at all times and without reasons. Future research may aim for baseline assessments prior to randomization, if ethically and practically attainable. Fifth, we did not assess intervention integrity. It is possible that therapists favored one version of YourSkills over the other, causing differences between conditions. Future research could assess intervention integrity by adding direct observations (e.g., videotapes of treatment sessions). Sixth, we used listwise deletion to deal with missing data, which could have biased the results. Yet, a robustness check using imputed data for the aggression measures yielded the same conclusions.

Our findings open up promising directions for future research. First, our study provides promising first indications that CBT with interactive virtual reality may

be more effective than care-as-usual for children with aggressive behavior problems; however, this is a first study and replication is needed. Second, building on the promising immediate post-intervention effects of CBT with virtual reality, it would be interesting to examine longer term effectiveness, as training effects can become more apparent when children have had more time to generalize learned skills to daily life (Lochman et al., 2015; McCart et al., 2006). Third, we included only boys in our study and findings can thus not be generalized to girls. Future research could examine whether girls with aggressive behavior problems benefit equally from adding virtual reality to interventions, or that adaptations need to be made in virtual reality scenarios. Fourth, it may be interesting to examine the mechanisms of change that may drive the decreases in aggression within children (Chorpita et al., 2005). For example, researchers could test if enhanced levels of emotional engagement and immersion in virtual reality predict larger decreases in aggression. Also, it may be relevant to test emotion regulation and social information processing as treatment mechanisms, especially as virtual reality may be a more effective tool to practice these skills. Fifth, investigating the cost-effectiveness of CBT with virtual reality may be a relevant next step, as this new technology comes along with extra costs for equipment, licenses to use the virtual reality, and training professionals (Lindner, 2021). This could be worth the investment if converging evidence shows that CBT with virtual reality is more effective than current treatments and may result in shorter treatments, less drop out, and lower costs for society on the long term (Geraets et al., 2021). Sixth, future research could examine therapist effects and for example investigate the influence of therapeutic alliance or years of experience on treatment outcomes (Karver et al., 2018).

In conclusion, we have found that it is likely that CBT with interactive virtual reality leads to larger decreases in children's aggressive behavior compared to care-as-usual. Compared to CBT with roleplays, results moderately favored virtual reality on four out of seven aggression measures, and clearly supported that virtual reality is likely to enhance children's emotional engagement and practice immersion, as well as treatment appreciation and perceived efficacy. Thus, interactive virtual reality seems a promising tool to enhance children's motivation during treatment and increase the effectiveness of CBT for children with aggressive behavior problems.

## FUNDING INFORMATION

This research was supported by a grant from the Netherlands Organization for Scientific Research to the last author (grant number 453-15-004/511).

## CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.


## DATA AVAILABILITY STATEMENT

Data and code for analyses are available at the Open Science Framework: <https://osf.io/dhkkq5/>. Most analyses presented here were preregistered in the Dutch Trial Register (NTR; <https://trialsearch.who.int/Trial2.aspx?TrialID=NL7959>).

## ORCID

Sophie C. Alsem  <https://orcid.org/0000-0002-6580-2114>

Anouk van Dijk  <https://orcid.org/0000-0002-6186-0181>

Esmée E. Verhulp  <https://orcid.org/0000-0002-5048-3480>

Tycho J. Dekkers  <https://orcid.org/0000-0001-8572-5606>

Bram O. De Castro  <https://orcid.org/0000-0001-5110-6153>

## REFERENCES

- Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*, *21*, 519–533. <https://doi.org/10.1007/BF00916317>
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Research center for children, youth, and families, University of Vermont.
- Alsem, S. C., Keulen, J., Verhulp, E. E., van Dijk, A., & de Castro, B. O. (2022). Capturing mechanisms of change: Weekly covariation in anger regulation, hostile intent attribution, and children's aggression. *Aggressive Behavior*, *48*, 232–240. <https://doi.org/10.1002/ab.22019>
- Alsem, S. C., van Dijk, A., Verhulp, E. E., & De Castro, B. O. (2021). Using virtual reality to treat aggressive behavior problems in children: A feasibility study. *Clinical Child Psychology and Psychiatry*, *26*(4), 1062–1075. <https://doi.org/10.1177/13591045211026160>
- Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health*, *3*, 1–31. <https://doi.org/10.2196/mental.4984>
- Berkout, O. V., Young, J. N., & Gross, A. M. (2011). Mean girls and bad boys: Recent research on gender differences in conduct disorder. *Aggression and Violent Behavior*, *16*, 503–511. <https://doi.org/10.1016/j.avb.2011.06.001>
- Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*, 303–316. <https://doi.org/10.1177/0049124189017003004>
- Burkey, M. D., Hosein, M., Morton, I., Purgato, M., Adi, A., Kurzrok, M., Kohrt, B. A., & Tol, W. A. (2018). Psychosocial interventions for disruptive behavior problems in children in low- and middle-income countries: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *59*, 982–993. <https://doi.org/10.1111/jcpp.12894>
- Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*, 566–579. <https://doi.org/10.1037/a0014565>
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, *7*, 5–20. <https://doi.org/10.1007/s11020-005-1962-6>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Earlbaum Associates.
- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837–844. <https://doi.org/10.1001/archpsyc.60.8.837>
- Crick, N. R., & Dodge, K. A. (1994). Review of social information processing in children's social adjustment. *Psychological Bulletin*, *115*, 74–101. <https://doi.org/10.1037/0033-2909.115.1.74>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*, 858–900. <https://doi.org/10.1037/a0038498>
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509. <https://doi.org/10.1037/0033-2909.131.4.483>
- de Mooij, B., Fekkes, M., Scholte, R. H. J., & Overbeek, G. (2020). Effective components of social skills training programs for children and adolescents in nonclinical samples: A multilevel meta-analysis. *Clinical Child and Family Psychology Review*, *23*, 250–264. <https://doi.org/10.1007/s10567-019-00308-x>
- DeVries, L. N., Hartung, C. M., & Golden, T. L. (2017). Negative halo effects in parent ratings of ADHD and ODD. *Journal of Psychopathology and Behavioral Assessment*, *39*, 179–188. <https://doi.org/10.1007/s10862-016-9560-z>
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). Deviant peer influences in intervention and public policy for youth. *Social Policy Report*, *20*, 1–20. <https://doi.org/10.1002/j.2379-3988.2006.tb00046.x>
- Fontaine, N., Carboneau, R., Vitaro, F., Barker, E. D., & Tremblay, R. E. (2009). Research review: A critical review of studies on the developmental trajectories of antisocial behavior in females. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *50*, 363–385. <https://doi.org/10.1111/j.1469-7610.2008.01949.x>
- Frick, P. J. (2012). Developmental pathways to conduct disorder: Implications for future directions in research, assessment, and treatment. *Journal of Clinical Child and Adolescent Psychology*, *41*, 378–389. <https://doi.org/10.1080/15374416.2012.664815>
- Geraets, C. N. W., van der Stouwe, E. C. D., Pot-Kolder, R., & Veling, W. (2021). Advances in immersive virtual reality interventions for mental disorders—A new reality? *Current Opinion in Psychology*, *41*, 40–45. <https://doi.org/10.1016/j.copsyc.2021.02.004>
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109–112. <https://doi.org/10.4103/2229-3485.83221>
- Hadley, W., Houck, C., Brown, L. K., Spitalnick, J. S., Ferrer, M., & Barker, D. (2019). Moving beyond role-play: Evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. *Journal of Pediatric Psychology*, *44*, 425–435. <https://doi.org/10.1093/jpepsy/jsy092>
- Hojtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*, 539–556. <https://doi.org/10.1037/met0000201>
- Hollis, C., Falconer, C. J., Martin, J. L., Whittington, C., Stockton, S., Glazebrook, C., & Davies, E. B. (2017). Annual research review: Digital health interventions for children and young people with mental health problems—A systematic and meta-review. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*, 474–503. <https://doi.org/10.1111/jcpp.12663>
- Hrabok, M., Brooks, B. L., Fay-McClymont, T. B., & Sherman, E. M. S. (2014). Wechsler intelligence scale for children-fourth edition (WISC-IV) short-form validity: A comparison study in pediatric epilepsy. *Child Neuropsychology*, *20*, 49–59. <https://doi.org/10.1080/09297049.2012.741225>
- Karver, M. S., De Nadai, A. S., Monahan, M., & Shirk, S. R. (2018). Meta-analysis of the prospective relation between alliance and outcome in child and adolescent psychotherapy. *Psychotherapy*, *55*, 341–355. <https://doi.org/10.1037/pst0000176>



- Kazdin, A. E. (2015). Treatment as usual and routine care in research and clinical practice. *Clinical Psychology Review*, 42, 168–178. <https://doi.org/10.1016/j.cpr.2015.08.006>
- Kort, W., Schittekatte, M., Dekker, P. H., Verhaeghe, P., Compaan, E. L., Bosmans, M., & Vermeir, G. (2005). *Wechsler intelligence scale for children—third edition—Nederlandstalige bewerking (WISC-III NL)*. Harcourt Test Publishers/Nederlands Instituut voor Psychologen.
- Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451–476. <https://doi.org/10.1007/s11292-005-3541-7>
- Lindner, P. (2021). Better, virtually: The past, present, and future of virtual reality cognitive behavior therapy. *International Journal of Cognitive Therapy*, 14, 23–46. <https://doi.org/10.1007/s41811-020-00090-7>
- Lindner, P., Miloff, A., Zetterlund, E., Reuterskiöld, L., Andersson, G., & Carlbring, P. (2019). Attitudes toward and familiarity with virtual reality therapy among practicing cognitive behavior therapists: A cross-sectional survey study in the era of consumer VR platforms. *Frontiers in Psychology*, 10, 1–10. <https://doi.org/10.3389/fpsyg.2019.00176>
- Lochman, J. E., Boxmeyer, C. L., Jones, S., Qu, L., Ewoldsen, D., & Nelson, W. M. (2017). Testing the feasibility of a briefer school-based preventive intervention with aggressive children: A hybrid intervention with face-to-face and internet components. *Journal of School Psychology*, 62, 33–50. <https://doi.org/10.1016/j.jsp.2017.03.010>
- Lochman, J. E., Boxmeyer, C. L., Kassing, F. L., Powell, N. P., & Stromeyer, S. L. (2019). Cognitive behavioral intervention for youth at risk for conduct problems: Future directions. *Journal of Clinical Child and Adolescent Psychology*, 48, 799–810. <https://doi.org/10.1080/15374416.2019.1567349>
- Lochman, J. E., Dishion, T. J., Powell, N. P., Boxmeyer, C. L., Qu, L., & Sallee, M. (2015). Evidence-based preventive intervention for preadolescent aggressive children: One-year outcomes following randomization to group versus individual delivery. *Journal of Consulting and Clinical Psychology*, 83, 728–735. <https://doi.org/10.1037/ccp0000030>
- Lochman, J. E., Kassing, F. L., & Sallee, M. (2017). Factors influencing intervention delivery and outcomes. In *The Wiley handbook of disruptive and impulse-control disorders* (pp. 485–500). John Wiley & Sons, Ltd.
- Lochman, J. E., & Matthys, W. (2017). *The Wiley handbook of disruptive and impulse-control disorders*. John Wiley & Sons, Ltd.
- Lochman, J. E., Wells, K. C., & Lenhart, L. A. (2008). *Coping power child group program*. Oxford University Press, Inc.
- Loeber, R., & Farrington, D. P. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology*, 12, 737–762. <https://doi.org/10.1017/S095457940004107>
- Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, 14, 545–555. <https://doi.org/10.1080/17405629.2016.1259614>
- Matthys, W., Maassen, G. H., Cuperus, J. M., & van Engeland, H. (2001). The assessment of the situational specificity of children's problem behaviour in peer-peer context. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42, 413–420. <https://doi.org/10.1017/S002196300100703X>
- McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: A meta-analysis. *Journal of Abnormal Child Psychology*, 34, 527–543. <https://doi.org/10.1007/s10802-006-9031-1>
- McConaughy, S. H., & Skiba, R. J. (1993). Comorbidity of externalizing and internalizing problems. *School Psychology Review*, 22, 421–436. <https://doi.org/10.1080/02796015.1993.12085664>
- Menting, A. T. A., de Castro, B. O., & Matthys, W. (2013). Effectiveness of the incredible years parent training to modify disruptive and pro-social child behavior: A meta-analytic review. *Clinical Psychology Review*, 33, 901–913. <https://doi.org/10.1016/j.cpr.2013.07.006>
- Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., & Kim, J. J. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: A randomized, controlled trial. *Psychiatry Research*, 189, 166–172. <https://doi.org/10.1016/j.psychres.2011.04.003>
- Polman, H., de Castro, B. O., de Thomaes, S., & van Aken, M. (2009). New directions in measuring reactive and proactive aggression: Validation of a teacher questionnaire. *Journal of Abnormal Child Psychology*, 37, 183–193. <https://doi.org/10.1007/s10802-008-9266-0>
- Rioux, C., & Little, T. D. (2020). Underused methods in developmental science to inform policy and practice. *Child Development Perspectives*, 14, 97–103. <https://doi.org/10.1111/cdep.12364>
- Silverstein, A. B. (1970). Reappraisal of the validity of WAIS, WISC, and WPPSI short forms. *Journal of Consulting and Clinical Psychology*, 34, 12–14. <https://doi.org/10.1037/h0028680>
- Sukhodolsky, D. G., Smith, S. D., McCauley, S. A., Ibrahim, K., & Piasecka, J. B. (2016). Behavioral interventions for anger, irritability, and aggression in children and adolescents. *Journal of Child and Adolescent Psychopharmacology*, 26, 58–64. <https://doi.org/10.1089/cap.2015.0120>
- Suveg, C., Southam-Gerow, M. A., & Goodman, K. L. (2007). The role of emotion theory and research in child therapy development. *Clinical Psychology: Science and Practice*, 14, 358–371. <https://doi.org/10.1111/j.1468-2850.2007.00096.x>
- Underwood, M. K. (2002). Sticks and stones and social exclusion: Aggression among girls and boys. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 533–548). Blackwell Publishers Ltd.
- UNESCO. (2012). International standard classification of education (ISCED 2011). *Prospects*, 5, 283–284. <https://doi.org/10.1007/BF02207511>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85, 842–860. <https://doi.org/10.1111/cdev.12169>
- van Manen, T. G. (2001). *Zelfcontrole [Self-Control]*. Bohn Stafleu Van Loghum.
- Verhoef, R. E. J., van Dijk, A., & de Castro, B. O. (2022). A dual-mode social-information-processing model to explain individual differences in children's aggressive behavior. *Clinical Psychological Science*, 10, 41–57. <https://doi.org/10.1177/21677026211016396>
- Verhoef, R. E. J., van Dijk, A., Verhulp, E. E., & de Castro, B. O. (2021). Interactive virtual reality assessment of aggressive social information processing in boys with behaviour problems: A pilot study. *Clinical Psychology and Psychotherapy*, 28, 489–499. <https://doi.org/10.1002/cpp.2620>
- Verhoef, R. E. J., Verhulp, E. E., van Dijk, A., & de Castro, B. O. (2021). Interactive virtual reality versus vignette-based assessment of children's aggressive social information processing. *Research on Child and Adolescent Psychopathology*, 50, 621–636. <https://doi.org/10.1007/s10802-021-00879-w>
- Weisz, J. R., & Kazdin, A. E. (2017). *Evidence-based psychotherapies for children and adolescents*. Guilford Publications.
- Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science*, 14, 216–237. <https://doi.org/10.1177/1745691618805436>
- White, I. R., Horton, N. J., Carpenter, J., & Pocock, S. J. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*, 342, d40. <https://doi.org/10.1136/bmj.d40>
- Wilson, S. J., & Lipsey, M. W. (2006). The effects of school-based social information processing interventions on aggressive behavior,

part II: Selected/indicated pull-out programs. *Campbell Systematic Reviews*, 2, 1–37. <https://doi.org/10.4073/csr.2006.6>

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. Update of a meta-analysis. *American Journal of Preventive Medicine*, 33, 130–143. <https://doi.org/10.1016/j.amepre.2007.04.011>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Alsem, S. C., van Dijk, A., Verhulp, E. E., Dekkers, T. J., & De Castro, B. O. (2023). Treating children's aggressive behavior problems using cognitive behavior therapy with virtual reality: A multicenter randomized controlled trial. *Child Development*, 94, e341–e361. <https://doi.org/10.1111/cdev.13966>